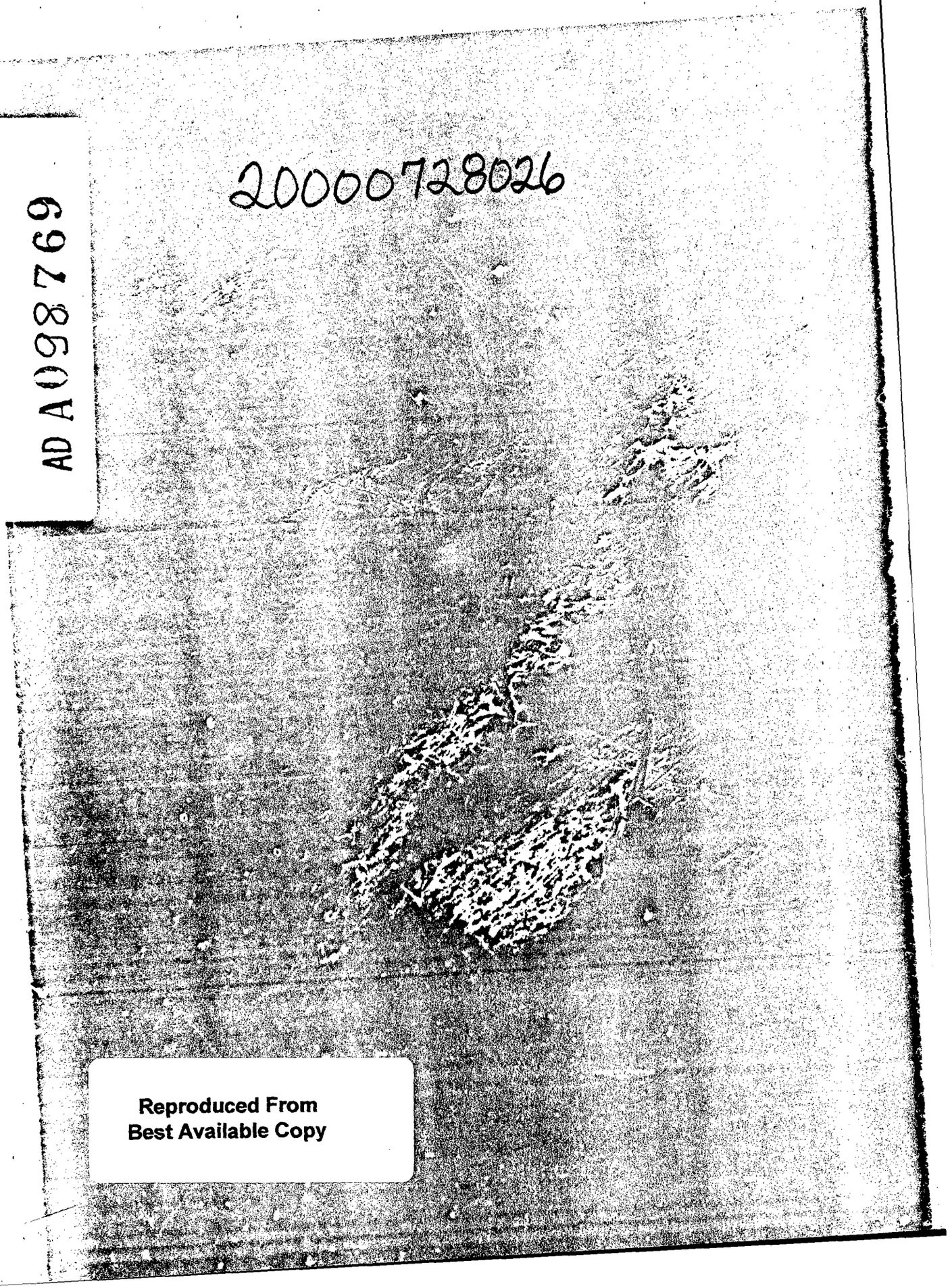
AD A098769

20000728026

Reproduced From
Best Available Copy



(12)

PROCEEDINGS OF THE ~~SEVENTH~~ SYMPOSIUM:

PSYCHOLOGY IN THE DEPARTMENT OF DEFENSE (7th)

16 April - 18 April 1980

(10) Victor H. Hansel
Richard I. Kosteeuk
William H. Clover
G. Howard Mickley
Carl Engvat

(11) June 80

(12) 662

MAY 12 1981

Department of Behavioral Sciences and Leadership

United States Air Force Academy

Colorado Springs, Colorado

(14) USAFA-TR-80-12

Symposium Chairman

Major Richard L. Hughes

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

EDITORS

Major Dickie A. Harris
Major William H. Clover
Major Richard I. Koeteeuw
Captain Carl Bryant
Captain G. Andrew Mickley

COMMITTEES FOR THE SEVENTH SYMPOSIUM

EXECUTIVE COMMITTEE

Major Richard L. Hughes, Chair
Colonel John W. Williams, Jr.
Lieutenant Colonel Jock C. H.Schwank
Lieutenant Colonel Gene A. Berry
Major Dickie A. Harris

PROGRAM COMMITTEE

Major Dickie A. Harris, Chair
Major William H. Clover
Major Richard I. Koeteeuw
Captain Carl Bryant
Captain G. Andrew Mickley

INVITATIONS AND ATTENDANCE COMMITTEE

Major Mark Nataupsky, Chair
Lieutenant Colonel Valentin W. Tirman, Jr.

SOCIAL, BILLETING & TRANSPORTATION COMMITTEE

Captain Stephan J. Pacheco, Chair
Captain Richard T. Smith

FACILITIES AND VISUAL AIDS COMMITTEE

Captain Lee J. Dahle, Chair
Captain Frank R. Wood
Technical Sergeant Frank C. Derry
Staff Sergeant Kenneth D. Fortenberry

PROTOCOL COMMITTEE

Major Robert W. Nordeman, Chair
Major Robert A. Gregory

This research report is presented as a competent treatment of the subject, worthy of publication. The United States Air Force Academy vouches for the quality of the research, without necessarily endorsing the opinions and conclusions of the authors.

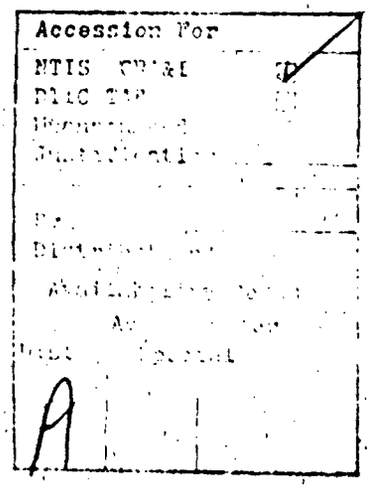
This report has been cleared for open publication and/or public release by the appropriate Office of Information in accordance with AFR 190-17 and DODD 5230.0. There is no objection to unlimited distribution of this report to the public at large, or by DDC to the National Technical Information Service.

This research report has been reviewed and is approved for publication.

M P Bacon

M. D. BACON, Colonel, USAF
Director of Research and
Continuing Education

Additional copies of this document are available through the National Technical Information Service, U.S. Department of Commerce, 5285 Port Royal Road, Springfield, VA 22151.



SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER USAFA TR-80-12	2. GOVT ACCESSION NO. AD-A098	3. RECIPIENT'S CATALOG NUMBER 769
4. TITLE (and Subtitle) Proceedings, Psychology in the Department of Defense, Seventh Annual Symposium		5. TYPE OF REPORT & PERIOD COVERED Proceedings 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Major Dickie Harris Major Richard Koeteeuw Major William Clover Capt G. Andrew Mickley Capt Carl Bryant		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Behavioral Sciences and Leadership		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS DFBL USAF Academy, CO 80840		12. REPORT DATE June 1980 13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Approved for Public Releaseand Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
Psychology, Symposium, Personnel, Training, Training Research Evaluation, Pilot Performance, Organizational Development, Job Enrichment, Women in the Military, Visual Display and Processing, Attrition, Stress Management, Human Factors Simulation, Leadership, Instructional System, Development, Radiation Biology, Clinical Psychology, Social Support Systems

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)
These printed proceedings include papers and presentations that deal with a wide range of research in psychology with emphasis on military issues.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

FORWARD

The opinions expressed in this volume are those of the individual authors and do not necessarily reflect official policy either of the Army, Navy, Air Force or of any government organization in which the author may be serving.

Trade names of materials or products of commercial or non-government organizations are cited only where essential to precision in describing research procedures or evaluation of results. Their use does not constitute official indorsement or approval of the use of such commercial hardware or software.

These proceedings are approved for public release, distribution unlimited, in accordance with paragraph IVA, AFR 80-45.

In publishing the Proceedings of the Seventh Symposium, the editors sought to facilitate an effective and timely dissemination of the technical information presented. The papers contained in this document were printed directly from unedited reproducible copy submitted by the authors who are solely responsible for their contents.

Acknowledgements

Cover, Graphics, Division of Audiovisual Services
Dean of Faculty, USAF Academy

Editorial Assistance: Joyce Wolford

OVERVIEW OF SYMPOSIUM SCHEDULE

Wednesday 16 APRIL 1980

Morning Activities	Event
	Registration
	Keynote Address Dr. Kenneth E. Clark
	Panel Discussions & Paper Session
	Panel: Conceptual Model for Evaluating the Performance of Service Academy Graduates
	Panel: Ground-Referenced Visual Orientation with Imaging Display
	Paper: Clinical Psychology
Afternoon Activities	
	Panel Discussions & Paper Sessions
	Paper Sessions
	Organizational Development
	Training and Training Research
	Performance Evaluation Testing
	Social Support Systems & Mission Effectiveness
	Panel: Community Mental Health Programs in DoD
	Panel: Mental Health Services to Basic & Technical School Trainees
	Panel: Institutionalizing Organizational Developmental Efforts: What Happens after the Initial Success?

THURSDAY, 17 APRIL

Morning Activities	Event
	Panel Discussions & Paper Sessions
	Panel: Selected Analyses of Air Force Organizational Assessment
	Panel: Air Combat Tactical Simulation Training
	Paper Sessions
	Evaluation
	Clinical Psychology
	Invited Papers
	Human Factors
	Personnel
	Invited Luncheon Address Major General G. E. Cooke
Afternoon Activities	
	Panel Discussions & Paper Sessions
	Panel: Pilot Screening Research
	Panel: Issues in the Professional Status of Psychologists
	Paper Sessions
	Personnel Evaluation
	Performance Evaluation
	Human Factors
	Personnel Issues in the Modern Army

FRIDAY 18 APRIL 1980

Morning Activities

Panel Discussion & Paper Sessions

Panel: Retention in the Air Force

Paper Sessions

Tutorial in the Design & Evaluation of Complex Man-Machine Systems

Leadership

Stress-Stress Management

Evaluation

PROCEEDINGS OF THE SEVENTH SYMPOSIUM ON PSYCHOLOGY
IN THE DEPARTMENT OF DEFENSE

WEDNESDAY, 16 APRIL 1980

KEYNOTE ADDRESS
Dr. Kenneth E. Clark

PAPER SESSION: Clinical Psychology
Chair: Sharon L. Slaughter

Psyop and the Role of Clinical Psychology
Richard W. Bloom

Examples of the Broadened Role of Psychology in Health Maintenance of Neuroendocrine Patients
Lawrence F. Wilson

Selected Factors Associated with Success in a Behavioral Weight Loss Program
Darrel Parisher

Inferences about Pain Patients
Rex A. Frank

PANEL DISCUSSION: Conceptual Model for Evaluating the Performance of Service Academy Graduates
Chair: Jack M. Hicks
Panelists: Lois DeFleur
Jerome Adams
Pat Harrison
Fran Hall

PANEL DISCUSSION: Ground Referenced Visual Orientation with Imaging Display
Chair: Stan Roscoe
Panelists: Jan C. Hull
Russell A. Benel
Nicholas M. Simonelli
Valerie J. Gawron
Joyce H. Iavecchi

PAPER SESSION: Organizational Development
Chair: Fred D. Harburg

Multidimensional Scaling Analysis of Unit Performance in Field Exercises
Steven M. Medlin and Paul Thompson

Relationship Supervisory and Job Variables to Organizational Effectiveness Variables
William H. Hendrix

The Impact of Organizational Conditions on Enlisted Attrition: Perceptions of Navy Human Resource Management Specialists
Gerry L. Wilcove and Ervin W. Curtis

Personal and Professional Identity
Frank R. Wood and David B. Porter

Cadet Opinions of the Best and Worst Aspects of the USAFA Academic Program
William H. Clover and David B. Porter

PAPER SESSION: Training and Training Research
Chair: Robert A. Gregory

Ages: From Development to Implementation
Earl S. Stein

Development of a Simulator Certification Methodology for Strategic Air Command
K. Ronald Laughery, Jr and George M. Houtman

Trainvice Analysis of Generic Electronic Maintenance Trainer Concepts
A. J. Josefowitz and J. W. Kochevar

Formulation and Evaluation of a Methodology for Predicting Required Hands-on Training Following Simulator Training
William R. Bickley

Simulator Training in the "Real World(s)": Some Issues and Empirical Answers!
Gerard M. Deignan and L. F. Cicchinelli

Design Requirements for a Planning Information System for Training and Personnel Technology RDT&E
Robert W. Stephenson and Matilde K. Stephenson

PANEL DISCUSSION: Community Mental Health Programs in DOD
Chair: V. P. Luchsinger
Panelists: Cecil Harris
John W. Williams, Jr
Harry Silsby
Gary Greenfield
Robert Anderson

PAPER SESSION: Social Support Systems and Mission Effectiveness
Chair: Cail I. Arnott

Family and Mission Conflicts: Their Impacts on Readiness
Cecile S. Landrum

The Relevancy of Counselor and Human Relations Skills Training for Air Force Personnel
Earl A. Glosser and Richard Percy

The Effects of Mobility on the Family and Service Member
Patricia C. Nida

PAPER SESSION: Training and Training Research
Chair: Edwin B. Griggs

Readability of Air Force Publications
Lydia R. Hooke

Meeting Reading Requirements for Military Service
Clay E. George and Adriana D. Barrera

Comparative Evaluation of CAI with Programmed Text and Lecture Modes of Instruction in Three Air Force Medical Training Courses
G. M. Deignan, B. R. Seager, M. Kimball, and N. S. Horowitz

Adaptations of Embedded Training
L. C. Percival

Comparison of Male vs Female Opinions Toward the Air Force Academy Academic Program
William H. Clover and David B. Porter

PAPER SESSION: Performance Evaluation Testing
Chair: Joel C. H. Schwank

Individual Differences in Multiple-Task Performance
• Diane L. Damos

Visual Search in Two-Region Clutter Displays
J. Wald, John R. Bloomfield, Laura A. Thompson, Deborah A. Smith

Brain Lateralization and Simple vs Integrated Task Performance
Linda D. Jackson, Richard L. Hughes, and Gene A. Berry

The Effects of Redundant Multidimensional Display Codes in a Complex Information Processing Task
R. E. Christ and G. M. Corso

• Manuscript not available for publication

Individual Differences and Laterality Effects in Dual Task Performance
C. D. Wickens, S. J. Mountford, Diane Sandry and W. Schreiner

PANEL DISCUSSION: Mental Health Services to Basic and Technical School Trainees
Chair: Thomas P. O'Hearn
Panelists: Thomas P. O'Hearn
James Jenkins
Hohn Stokes
Henry Lagarde

PANEL DISCUSSION: Institutionalizing Organizational Development Efforts: What Happens after the Initial Success?
Chair: Vincent P. Luchsinger
Panelists: William Rosenbach
Michael Pavich
William Clover
Denis Umstot

THURSDAY, 17 APRIL 1980

PAPER SESSION: Evaluation
Chair: Mark Nataupsky

The Human Resources Test and Evaluation System (HRTES) Its Concept and Design
D. Kaplan and H. Crooks

Evaluation of an Image Quality Assessment Technique Based on Magnification
Gilbert G. Kuperman

A Human Operator Model for Command and Control Evaluations
Charles Jorgensen

The Role of the Operator in Designing for Maintainability
James W. Kochevar and J. F. Brock

Improvements in U. S. Army Rifle Marksmanship Based upon Recent Marksmanship Research
Seward Smith, Thomas J. Thompson, John C. Morey and Arthur D. Osborne

PAPER SESSION: Clinical Psychology
Chair: Carl Bryant

A Self-Concept Analysis of Cross-Cultural Marriage Variables
Jaime I. Albornoz

Social and Familial Role Perceptions as Correlates of Self-Concept and Locus of Control in Adolescents
Robert J. Drummond and Walter G. McIntire

Differences in Family Coalitions and Hierarchies between Normals and Neurotics
Vance Shepperson

PAPER SESSION: Invited Papers
Chair: Mickey R. Dansby

John Henry Syndrome: Man's Conflict with Automation
Duncan L. Dieterly

Behavioral Contributions to Strategic Planning
V. P. Luchsinger and Otis Bryan

The Risks of the Cuban Missile Crisis: Controlling the Analysts
Richard Ned Lebow

The Variables that Predict User Satisfaction with Research: A Multiple Regression Approach
William C. A. Stewart, Thomas J. O'Connor

Antihistamines Eliminate Sex Differences in Post Irradiation Performance
G. A. Mickley

PANEL DISCUSSION: Air Combat Tactical Simulation Training
Chair: Ronald G. Hughes
Panelists: David Bessemer
Dirk Prather
Jack Haile

PAPER SESSION: Human Factors
Chair: William G. Buchta

A Functional Metric for Description of Visual Stimuli
R. A. Benel

New Methodologies for Target Acquisition Modeling
*G. E. Corrick

Cognitive Maps and the Subjective Magnitude of Large-Scale Distances
Edward J. Rinalducci and Douglas A. Hardwick

Visual Information and Sensitivity to Loss of Altitude
*S. J. Mangold, R. Warren, D. H. Owen, R. S. Jensen, and J. Boykin

*Manuscript not available for publication

Eye Accommodation, Personality, and Autonomic Balance
Valerie J. Gawron

PAPER SESSION: Personnel
Chair: William H. Clover

The USAF Academy How-to-Study Program: Past, Present and Future
Mark Nataupsky, John Bermudez, Valentin Tirman, and Nita Huelf

Individual Choice, Personal Affect, and Personal Adjustment of Women who have Resigned from West Point
Jerome Adams, Howard T. Prince II, and Francine Hall

Male and Female Differences in Army Career Branch Selections at West Point
Janice Calhoon and Bret Dalton

Cadet Interests in Flying Specific Aircraft
J. M. Koonce

An Examination of the Impact of Environmental Stressors on Cadets during Summer Training at West Point
Douglas McCord

PANEL DISCUSSION: Selected Analyses of Air Force Organizational Assessment
Chair: Lawrence O. Short
Panelists: Fulton D. Hayes
David Wilkerson
Jerry R. Pfeifer
Lawrence O. Short
Charles R. Spicher
Kenneth L. Hamilton
Glenn E. Boren

PAPER SESSION: Personnel Evaluation
Chair: David B. Porter

Conceptual Model of a Merit Pay Appraisal System for Air Force Civilian Employees
John A. Guerrieri, Douglas Cowan, and Bart Vitola

Initial Development of a Job Performance Evaluation System (JPES) for Air Force Civilian Employees
Thomas W. Watson, John A. Guerrieri and James W. Earles

Rater Accuracy in Predictions of Ability
William H. Cummings

The Effect on Leniency of Justifying Performance Ratings to a Superior Officer
Phillip J. Decker and Edwin T. Cornelius III

Type I and Type II Errors in Performance Evaluation
Frederick H. Steinheiser

Validating the Role Reaction Model for Predicting Superior-Subordinate Relationships
V. Ann Sweney and Arthur B. Sweney

PAPER SESSION: Performance Evaluation
Chair: Thomas M. McCloy

Time Estimation: Repeated Measures Testing and Drug Effects
Alvah C. Bittner, Jr., Robert S. Kennedy, and Michael E. McCauley

Performance Evaluation Tests for Environmental Research (PETER): Code Substitution Test
Ross L. Pepper, Robert S. Kennedy and Alvah C. Bittner, Jr.

Performance Evaluation Tests for Environmental Research (PETER): Interference Susceptibility Test (IST)
Michele Krause and Robert S. Kennedy

Video Games and Convergence or Divergence with Practice
Marshall B. Jones, Robert S. Kennedy and Alvah C. Bittner, Jr.

The Effect of Cognitive Style on Performance of a Critical Tracking Task
T. M. Mabry, D. A. Harris, and G. A. Berry

PANEL DISCUSSION: Pilot Screening Research
Chair: David L. Payne
Panelists: Jefferson M. Koonce
E. Craig Campbell
Harald Jensen

PAPER SESSION: Human Factors
Chair: G. Andrew Mickley

The Pilots' Night Training Device--An Answer to the Army's Night Terrain Flight Training Enigma?
Colin D. Ciley and H. F. Allnutt

Glideslope Descent Rate Cuing to Aid Carrier Landings
Gavan Lintern

Simulation Requirements for Helicopter Landing
G. L. Ricard, R. V. Parrish, B. R. Ashworth and M. D. Wells

Pilot-Computer Interfaces for Complex Aerial Navigation: Simulations Using a Microprocessor
D. B. Beringer

How Should You Talk to a Semi-Intelligent Machine?
J. K. Arima and Anthony G. Quartano

Use of Imagery-Based Training Materials for Learning Procedural Tasks
Gary G. Miller

PAPER SESSION: Personnel Issues in the Modern Army
Chair: Stephan J. Pacheco

Non-Conforming Behavior in the Early Volunteer Army
Larry J. Crisler

The Volunteer Army and Initial Adjustment Problems
E. R. Worthington and James Warren

Female Adjustment in the Volunteer Army
E. R. Worthington

Some Army Psychologists Remain in the Service - Why?
A. David Mangelsdorff

PANEL DISCUSSION: Issues in the Professional Status of Psychologists
Chair: John M. Bermudez
Panelists: Jefferson M. Koonce
Donald Trahan
Dale Mann
Anthony M. Ricci

FRIDAY, 18 APRIL 1980

Tutorial in the Design and Evaluation of Complex Man-Machine Systems
Participants: Terry Emerson
Richard Moss

PAPER SESSION: Leadership
Chair: Linda D. Jackson

A New Approach to Undergraduate Leadership Education at the Air Force Academy: Combining Psychology with History
L. F. Silverman and Fred D. Harburg

Is Flying Enough? Pilot's Interpersonal Skills Training
Judith Richter

Leader Effectiveness as a Function of Leader's Sex and Leadership Style
Mickey R. Dansby

Gender and Organizational Leadership: A Comparison of Two Models
Thomas McCloy

Prediction of Military Order of Merit Using Assessment Center Methodology
Hal W. Hendrick

Perceptions of Psychological Influence: The Role of Person-Environment Fit in Subordinate Reactions to Leadership
Lawrence R. James and S. B. Sells

PANEL DISCUSSION: Retention in the Air Force
Chair: Robert A. Gregory
Panelists: Robert H. Smith
Willibrord T. Silva
James E. Cochran
Warren C. Patton
William E. Rosenbach

PAPER SESSION: Stress-Stress Management
Chair: Lee J. Dahle

Psychopharmacological Intervention in Bereavement and Grief
Victor M. Solomon

The Impact of Stress in the Air Combat Environment
Dana R. Ideen and Jeffrey E. Kantor

Interpersonal Stress and the Utilization of Student Aptitude
Earl H. Potter

The Differential Effectiveness of Specific Treatments on the Worry-Emotionality Components of Test Anxiety
Kermit E. Parker, Jr.

Test Anxiety and Psysiological Response to Evaluation Stress
Scott W. Carrothers and Richard L. Hughes

PAPER SESSION: Evaluation
Chair: Richard I. Koeteeuw

Effect of Alternate Instrument Displays on Novice Pilots
Ross A. Cannon and Jock C. H. Schwank

Effects of Presentation - and Test - Trial Training on Motor Acquisition and Retention
Joseph D. Hagman

USAFA Male and Female Cadet Comparisons in the Prediction of Psychomotor and Basic Flight Skills
J. M. Koonce and G. A. Berry

Gender Differences in Aircraft Simulator Performance
Thomas McCloy

The Use of Personality Characteristics as Predictors of Psychomotor Performance
G. A. Berry, D. A. Harris, and J. M. Koonce

Training Criteria: The Missing Link in Flight Simulation Research?
Mark Nataupsky, J. M. Bermudez, and V. W. Tirman

Keynote Address

Kenneth E. Clark, University of Rochester

I'm delighted to be invited to this annual conference of psychologists in the Department of Defense, and to be able to stay through the session and to hear the variety of important things that psychologists are doing in the Army, the Navy, the Air Force, and the Marine Corps. It is gratifying to note today how much we are appreciated, and how much we are sought after to provide answers to a variety of questions. The fact that the Department of Defense would support a conference of this sort, that there would be so many persons to attend, and that the reports we shall hear will be such a small proportion of a vast amount of work being conducted would not have been predicted in earlier days. Furthermore, to have all of us identified as behavioral scientists and not as corporals, captains, and majors in the sanitary corps, makes a monumental difference in who listens to the work of psychologists, and how much things have changed since World War I.

In a sense Psychology has grown along with all of the other sciences utilized by organizations and by various sectors of government. But in a sense we are different, for the growth in the attention to and the use of behavioral scientists has been beyond anyone's predictions. As recently as twenty years ago, one would not have predicted this. In the report of the Life Sciences Panel of the President's Science Advisory Committee on strengthening the behavioral sciences, which was issued by the White House, in the agency for international development and in the evaluation of new social programs that were launched, the outlining was rather explicit. There is no mention at all of the way in which the United States could improve its defense establishment through the use of behavioral sciences.

In a sense we would like to feel very proud of that, and to thank our friends for their cleverness and their insight in seeing that we had something to contribute. However, it is a little bit out of control in a certain sense. That is all of the use of the behavioral science data and procedures, and all of the contributions that we would call the contributions of psychological research are not contributed by scientists, and our names are taken in vain frequently, our findings distorted, and sometimes objected to. Another factor associated with this, is that those who are doing this work do not have adequate background,and borrow our techniques improperly.

It is also true that all of the use of psychological science cannot be called science. That sounds the same as what I just said -- I do not intend it to be the same. The way in which psychological research in the identification of individual differences is interpreted, the way in which answers are given to questions, are frequently not ways that we would call deeply rooted in the scientific method. All of those that have been called upon to give advice to the agencies for which you work know that sometimes you're going as much by the seat of the pants as anything else, and that you ought to have been called a management consultant, or a systems designer, rather than a scientist.

I do not intend however, to criticize those who want to see us do good things for the society. I am strongly in favor of that, and would

argue that while many of things we are asked to do are beyond our capabilities at the moment, we should nonetheless be asked to do them. And if the request is beyond our capabilities, we must strengthen and increase our own resources in order that we may be better able to meet the demands in the future. My first theme in this talk therefore is that it is essential that we respond to those who ask us to help, and to respond in a constructive manner. Lest you think in terms of last week's memo from your superiors, let me remind you how wide is the variety of issues on which psychology is expected to make contributions. How do you help people quit smoking? What programs would you suggest? Which ones work? Under what circumstances do they work? Are we talking about a habit that can be unlearned, or an addiction that needs to be treated? Is smoking a medical problem or is it a behavioral problem? How can we help people to lose weight? Is obesity a physical problem or a behavioral problem? How do we get people to get off drugs? Again, is that an addiction that requires physical treatment, with drugs and other medical procedures, or is it a behavioral problem? How do we reduce violence in the family? Is wife-beating inherited or is it acquired? How do we reduce schizophrenia? Is schizophrenia produced by schizophrenogenic mothers, or by an inherited characteristic or some systemic imbalance? How do we teach students to study? How do we debrief a "Moonie"?

How do we set up a national program so that people want to serve in the Army? Or the Navy? Or the Marine Corps? Or the Air Force? What are the issues that psychologists can make a contribution to as we discuss a draft as against a volunteer service? Or are these questions that we should leave to the economists and to the managers, and assume that we have no contribution?

How do you keep Air Force pilots in the Air Force? Are there any such things as "pilot errors" or does every error by a pilot indicate something wrong with the system that requires correcting? If there are "pilot errors", how do we reduce them?

How do we keep skill levels up for tank crews who have learned to zero in on an enemy target and fire within seven seconds? We notice that the forgetting curve has certain unhappy characteristics. How can we offset that forgetting?

How can we reduce accidents on the highway? How can we reduce drinking and driving? Are safety characteristics of automobiles, and characteristics of highways more important than the characteristics of drivers? Stated more appropriately; How much of the variance in automobile accident rates is associated with characteristics of cars, characteristics of highways, characteristics of weather, characteristics of traffic flow, and characteristics of traits of drivers and conditions of drivers?

It is obvious that this list can go on forever. After a while it gets kind of boring. Let me mention several things that suggest that the importance of the psychologist as a student of human behavior and the needs for psychologists as consultants and managers of social systems will increase in the future rather than decrease, whether we call the person doing that work a psychologist or not.

Training of riflemen in the Army requires each rifleman to fire at least three hundred rounds on the range. Three hundred rounds has been the standard for a long time. Yet the curve of learning for riflemen on the

range shows that after fifty rounds there is no further improvement, even though the level of accuracy has not reached the level one had intended. What is the purpose of the additional two hundred and fifty rounds? Is this a tradition that was established from the folklore that we ought to stamp out, or is there some other role to be played by that extra firing? Studies of riflemen in combat suggest that it is more important that they fire frequently than that they fire with great accuracy at a pinpoint target. What implications does this have for the training of riflemen? Persons who have fired a great deal before coming into the Army do not necessarily become the Army's best riflemen. If what we were interested in then is accuracy and rapid learning, women who have never fired a weapon before can be taught to fire it better than most persons who have used such weapons continuously. This is through use of the Army weaponaire which simulates the firing, and prevents the immediate establishment of anticipatory responses to the recoil and to the noise of the shot by building these two gradually into the system. When a person so trained goes to the range, the transfer is near perfect for the assignment given. However, the assignment is to aim at pop-up targets that are very similar to the situation in which the soldier trained.

I'm sure that this illustration is known to some of you, and that perhaps some of my details are wrong. But the principles are known to all of you, and involve large portions of basic psychology including some application of the early work of John B. Watson on the reduction of fear, and desensitization to fear, the work on basic learning theories about the nature of the learning curve and the way in which it can be influenced, important early work that most of us have forgotten about transfer of training, and the conditions under which it occurs, interference effects in learning, which we all have studied as part of our graduate work in psychology. We could also add a couple additional paragraphs to this story, commenting that most objects that a soldier fires on in combat are moving across the field of vision, and are not pop-up targets, and that this sort of training would not necessarily apply to that. That sort of evidence we already know, and so we emphasize rapid fire in addition to precision target shooting.

This illustration leads me to my second major point of the morning, and that is that applied psychology and applied behavioral science is not a separate field from basic science in psychology, and that in fact the two are very hard to distinguish. We cannot say in psychology what is sometimes said in the physical sciences, that the engineers have run out of basic science on the shelf in certain of their problems, and that they need a development of further basic science. A large portion of psychology has grown out of applied psychology. World War I gave a substantial impetus to the development of testing in psychology, because of the great attention that the use of Army-Alpha and Army-Beta produced during World War I. What is not so much noticed is that a large number of other segments of society, including questionnaire techniques, measurement of personality or measurement of motivation for work, got their start from groups that were involved in the World War I effort and who saw the needs for further developments and proceeded with them. The role of the psychologist in the use of personality tests has today gone far beyond that where the role of the psychologist as a clinical practitioner is based only to a

very modest degree on that person's diagnostic capabilities. If one looked to the domains of basic research in psychology that led to that application of the services of psychology, one would look in vain. What writings there are have been writings that were in the main conjectural, such as Abe Maslow's interesting and informative analyses of need hierarchies, and have followed the development of the practitioner as a helper in the community rather than leading it.

There is much to be said for the same thing operating in the domains of psychology with which we are all concerned today. To a certain extent th needs of the society have led us into doing things we did not know how to do. Psychological science gives us guides on the ways to proceed in the analyses of problems of the sort with which we must cope, but on the main we are often on our own.

When a command asks for some guidance in improving the quality of the performance of crews, so as to increase combat effectiveness, the psychologist has a way of looking at that, he has a way of analyzing that problem that is distinctive. The background of all of us includes methods of quantitative analysis which take into account ways of determining changes in behavior, ways of assigning changes to appropriate circumstances, and ways of determining those more or less invariant characteristics of individual persons which will influence the way in which those changes occur. We also have learned certain ways to go about the analysis of a problem. A good illustration of that is the analysis made by a distinguished Boeing psychologist on trying to understand why it was that 727's were crashing. You will remember that for a long time there was a discussion of the sink-rate of the 727 being greater than that of other planes. Laboratory studies at Boeing indicated that in the particular instances involved that was not the critical factor. It turned out that each of those crashes involved landing a plane at dusk, and that at dusk an airfield, because it is darker than the surround, appears to be lower down than the rest of the terrain under particular circumstances. The most critical circumstances is the lights of the city be beyond the plane at the time that the pilot is landing. Under such circumstances the pilot assumes that he is higher than he actually is, and thus drives the plane into the ground. This illusion had existed always, of course, but the 727 at full throttle drives the plane down until the lift makes it possible for the pilot to raise the nose of the plane. That driving down, when the plane is at too low a level, produces a disaster. What is now going on is that co-pilots are required to read off the altimeter under these conditions to assure that the pilot does not gauge incorrectly his altitude as he comes in for a visual landing. Here again we see an illustration of a problem that includes classical aspects of psychology. We have always been fascinated with illusions. It also incorporates a way of controlling behavior that places reliance on more than visual cues. We learned many times in the laboratory that using more than one sense modality will promote learning and will promote accuracy. That the problem was properly analyzed was undoubtedly due in part to some sort of creative insight on the part of the investigator, who put together knowledge that he already had in an unusual way.

We psychologists possess all kinds of peculiar items of knowledge, that we can bring to bear in this particular way. Most people do not

credit psychology with having as an extensive knowledge base that it does. Everybody assumes that psychology more or less applies the folklore to ordinary everyday problems, and that anybody can do it. We need to disabuse people of this. We know a lot of things that other people do not know. We should take pride in that knowledge and apply it when we can.

Most of you undoubtedly find that your superiors are not sure that you know anything more than they do. I think it's useful to collect items of information about psychology that people would not believe, would guess wrong, in the absence of knowledge, so that you may use such items to illustrate that you know something that other people don't. Besides it's a fun parlor game.

Let me try a couple out on you? How long can a healthy athletic man hang by his hands from a bar, with his feet off the ground? If any of you think you can do it for more than three minutes, try it. That is, assuming that all of you identify yourselves as healthy young men or women. If you were shown one hundred pictures clipped out of high school yearbooks, and half of them were from your own high school yearbook, how many of them would you be able to say that you had seen before, and be correct? The answer to that question is psychological, it's based on the work of Roger Shepard, who is an outstanding psychologist of perception, and he will predict that you will be more than 95% accurate, regardless of how long ago it has been that you graduated from high school. Suppose you take one hundred random snapshots that you had never seen before, and looked through them, just flipping them through one every two seconds or so, and then you shuffle those with another hundred snapshots that you had never seen before and now you try to sort out the ones you had seen before from the ones you hadn't seen before. What would your accuracy rate be? Again, about 95, 98% accurate. So people who say, "I can never remember a face," are wrong. You can remember faces. The problem is remembering names. Ben Underwood has done a lot of study about memorizing nonsense syllables and demonstrates that you can remember nonsense syllables and recognize them. The problem is pairing the two and people who say they cannot remember names can remember the faces, may remember the names, may never have learned the names in the first place, but even so have their main problem in pairing the name and the face. That sounds trivial to you. That sort of analysis has never occurred to lots of people, and so they go on never learning to pair names and faces because they had never analyzed the problem. If you want rapid association of names and faces, it is necessary that the name be familiar to you before you see the face. As we all know. Why are mothers more interested in their children getting acquainted with their cousins than the cousins are in getting acquainted with each other? If that's not characteristic in your family, I can report that it is characteristic in mine. The answer is that mothers are more closely related to their nieces and nephews than cousins are to each other, and that people tend to give highest priority in association with family to those persons with whom they are most closely related by genotype. This is one finding of the sociobiologists that is not controversial. I'll go to the next statement at this point.

We know some things that we find it rather unpopular to say. We know that in World War II, before a program of civil rights to improve the lot

of blacks in this society, that the average black scored one full standard deviation between whites in a population of 9,000,000 recruits tested during that period. We know today that that gap in intelligence test scores has not reduced very much, in spite of more money being spent on education, and more educational opportunities presumably opened for blacks. A lot of people do not want to believe that. We also know that these test scores are closely associated with performance on the job, and so that the differences mean something. We also know that investing more in schooling does not produce any changes in these scores that is appreciable. We also know that if the tests are rewritten so as to be culturally unbiased and insofar as possible non-verbal, the difference remains.

We also know some things about learning that are unpopular for us to say. The first is that many traits that we consider learned show up to appear in identical form in identical twins reared apart. We also know that those behaviors that receive reinforcement in any form tend to habituate, tend to repeat, and that many persons with impossible children deserve them, in fact train them that way, and that families in which the pet is in charge are families which ordered it that way by their own behavior. When we observe the pathologies of the society today, and remember a time when these pathologies did not exist, we must take note of the fact that we produced those pathologies, and that we must have wanted them as part of what we bought as we developed certain social programs. We know, for example, that a program of influence on a group will work in a small town but will not work in a big city. Yet we pay people to live in big cities by the nature of our welfare programs.

We also know that behavior that has been evidenced in the past is the best prediction for behavior in the future, everything else being in about the same state at each time. So a person who is lousy in one job will probably be lousy in another. We often hope that that is not true and promote people to get them out of a job that they are doing poorly. I really do not believe in the Peter Principle. I do not believe that people are promoted beyond their level of competence. I think many people are promoted because they're incompetent. As a result, our leadership in many settings is much less than what we want. It is interesting that this phenomenon does not occur in the highest ranks in the military, and that among the persons who do achieve major leadership positions, that they have qualities that can be greatly admired. I greatly admire the quality of the very top brass in the Army, because they insist that their newly designated Brigadier Generals go through some sort of education that will help them understand a little bit better the way in which they work with human beings. They get sent for a leadership development program at the Center for Creative Leadership in Greensboro, North Carolina.

An interesting aspect of the various things that psychologists know that others do not has to do with drug addicts. Those of you who have not read Nick Cumming's article in the December issue of The American Psychologist ought to do so. He indicates why it is that drug addicts prefer psychologists or therapists rather than psychiatrists. Psychiatrists treat drug addiction with drugs, psychologists treat the problems that led to the addiction, and develop a program of re-education that leads not to addiction on something else but to non-addiction.

My point in the illustrations I have given is to emphasize the fact that psychology and applied psychology cannot be distinguished from each other. That psychology and its applications are all of a piece. I've read through the titles of the papers for this program, and again feel that the topics being addressed by psychologists in the Department of Defense are as central to the domains of knowledge of all psychology as are the research activities being conducted in laboratories in our purist basic research institutions. In fact, it may be that the future of psychology rests more with the topics that are considered in this conference than in the topics that would be covered in meetings of let us say, the Eastern Psychological Association. I do not find this hard to understand, for I have noted that the most able of psychologists are motivated to understand all of human behavior, and that they too make no distinction between basic research and applied. Arthur Melton would certainly qualify as one of the leaders of experimental psychology, yet his contributions to the Air Force in research activities are well known. Harry Harlow is best known for his issues of mother love and the monkey, but he helped the Army establish an important research program. During World War I, Yerkes was the leader of the psychological group. And yet today, he would be considered not an applied psychologist at all. I could continue this list endlessly but I do not need to. The issue of whether applied psychology is part of psychology is a dead issue. It was not dead before World War II, for then the applied psychologists felt so neglected within the American Psychological Association that they formed their own association -- the American Association of Applied Psychology. At the end of World War II, the two groups came back for a happy marriage, with a structure of the APA modified to allow both to co-exist in comfort.

Today the issue in psychology is not an issue of basic versus applied, although there are problems about whether all of psychology belongs together. We have large numbers of students around the country majoring in psychology as an undergraduate major. I applaud that. I think it is to the undergraduate student's advantage not to make a choice of career until the student has had an opportunity to try out a large number of ideas. Making a decision to be a psychologist is making no decision at all, for the psychologist is as much at home in the political arena, in the business arena, in the hospital area, in the halls of academe, as he is in the psychological laboratory. So in a sense, making a choice of major a psychologist leaves open a wide variety of options.

It is that great diversity of psychologists that is giving the American Psychological Association problems in terms of its organization. Those psychologists who spend their time developing materials for publication, and who teach others, find little interest in those persons who spend a large proportion of their time seeing individual clients in private practice, and who are seeking to find ways to keep only the most competent providing such services. The psychologist in a military setting finds it very easy to make a decision about which group to join there because the military psychologist uses all of psychology and finds himself in constant association with the basic literature in psychology.

This same issue of diversity crops up when we try to figure out how to organize journals. It is an unhappy fact that for those people who pay for the journals, that population reads only 4% of the pages they pay for.

The journals on library shelves in graduate departments get more readership, for graduate students read a lot of dismal stuff to find out what's been going on in the past. For those of us who want to know what is going on now, we come to meetings like this and write to the persons we know are engaging in such research, rather than waiting the two years until an article about that work shows up in the published literature, or three years when we can find it organized in the psychological abstracts. Those of us who don't care about keeping up, can wait even longer and read the annual reviews of psychology and find out what the field was like three, four or five years ago.

Another interesting feature of the attractiveness of the skills of the psychologist is that other people are moving in on the field. Whereas at one time a psychologist studied vocational choice and developed interest inventories, and counseled persons about educational and career decisions, the sociologists have entered the field using occupations as a prime illustration of problems in stratification, and the economists have invaded the field because they have been called upon to suggest ways in which the human resources of the country could be used more effectively through control of aspects of the economy. In the past we have always thought that our nearest neighbors were first philosophy, although that may be more of a grandparentage than a neighbor, sociology, biology, and psychiatry. It may be in the future that our nearest neighbors will be political science and economics, as psychologists move more and more into those domains in which they influence the societal decisions about dealing with issues of the times.

I wish I could applaud this move. I find a certain discomfort in the developments I see in terms of the ways in which people seek help for the problems that exist. Let me illustrate -- the economist's major error is that he engages in too much aggregation, and makes decisions about groups in terms of intergroup variability rather than in terms of the variance within groups. As we all know, you can end up with exactly the wrong conclusion, if you correlate statistics rather than correlating the measures for individual persons. The general error of economists is the postulating of economic man. With the development of macroeconomics, the story is different but the underlying concept is the same, namely that everybody reacts the same to a buck. Stated in polite economic society as the effect of market conditions. The managers we work with tend to solve their problems by instructing those under them to tell the guy what to do and presume that he will do it. The manager's subordinate responds, "I did tell him and it didn't work." In our society today we hear some thunder from the right which says that nobody wants to work anymore and some chatter from the left that says we must provide for the sake of the old and the poor. The psychologist has something to say about both of those, for one of them assumes that nobody is motivated and the other assumes that everyone is altruistic. We are invited to engage in studies that will have some effect on law and order, with the assumption that we must find that form of punishment in single individuals that will inhibit the behavior of thousands. We are considered as opponents of civil rights, and have people telling us that we should quit making their job harder by saying that some people are dumb or incapable or that certain groups are inferior to certain other groups.

So there is a great deal for us psychologists to do in the society of the future. We are being asked to do it. You as psychologists working in the military are asked to do many things I encourage you to try to do them insofar as you can, and to indicate the limits of your capability. Decision making is often required rather than research studies. We cannot use our scientific capabilities as a substitute for making hard decisions. We can provide a better data base on which to make such decisions.

I have read with interest, and I hope you all have read, the book by Peter Watson, War on the Mind; the Military Uses and Abuses of Psychology. I started that book and almost threw it out the window after I had read the first few pages. I recommend that you continue reading because it is an excellent summary of what psychology can do, and not a diatribe against the use of psychology in the military service. Watson does raise some issues that need to be raised, but I believe that they are raised needlessly as in a context that assumes war is such hell that one should not contribute anything to turning it into only a purgatory.

The military has always capitalized on the scientific capability of the times. That exploitation of science has in turn led to the development of more science. Those of you who read history of science more deeply than I can, could cite many instances. I know of enough to be confident that the military effort of the time will contribute greatly to the development of psychology.

In fact, I have more confidence in the research influences of psychologists in other behavioral sciences in the military setting than I do for similar persons in the civilian setting. Today we are seeing a great know-nothing movement. Large numbers of our young people turn aside from knowledge and accept faith and belief in drugs and the moods of the mind and the myths of the spirit as a substitute for the problems of living and the necessity of making decisions about one's life. I deplore this, but I see more of it coming. I believe that we will see a substantial increase in know-nothingness in the next generation. That means that psychological techniques will continue to be suspect. When we learn ways to make prisoners and mental patients more content and more satisfied with their state, we will also be accused of having violated their civil rights. When we aim to analyze properly the factors that lead to very slow improvement in the status of blacks and minorities in employment and in our educational systems, we will be accused of being racist. In the military there will be a little less opportunity for this to occur, although the virus is there also. There will be much less attention to these facts just as there is less attention in business, because the ultimate criterion of performance will be performance. True combat effectiveness must be maintained. We cannot merely assume that if we have the right mix of persons of a variety of social and sexual origins that it will automatically occur. The quality of the object that comes off the end of the assembly line can be determined objectively, and the sources of variance determined so that in those instances where there is an objective criterion, a hard measure against which to test a given system of social influence, we are more likely to have truth. And it is the truth that it is important for us to attain. We cannot organize a society on a bundle of myths. If we do we are as certain to decline as prior civilizations declined. That decline will be more rapid if there isn't any other social system challenging our own. During

the last sixty years, the primary contender for a social system other than our democratic capitalistic system with an open society and with human beings generally free has been the communist system. Today we have a third contender, a religious system. Let us not assume that the Islamic state in Iran is somehow or other going to fall of its own weight and destroy itself because it is so senseless and irrational. That system is the result of a systematic conclusion about how to have the best life by a large number of persons, apparently a majority of all the persons in Iran. We should watch that development carefully, for it may be that it will lead to events that would make the Scope's trial in Tennessee appear trivial and a minor incident in a society.

I would like to assign to psychology and psychologists a higher morality than to other segments of the society. All of us came into psychology because we believe that by the work of psychologists and the knowledge of psychology, we could improve society and help solve human problems. I believe that most of us are committed to that. Given the choice between making Generals happy and the troops happy, we would generally prefer that the troops be happy. We would think that first because of issues of the common good, but also because we believe that there is a more effective fighting army when the troops are happy. Our problem is to assure that both occur. Sometimes we will worry about the priorities that we give to the work that we do. I would argue that the work is important, that is is central to the development of better psychological science in the future, and that it is central to our development of the good life in the country. I hope your conference is a great success for all participants, and that we learn a lot, and that we may progress toward our common goals.

PSYOP and the Role of Clinical Psychology

Richard W. Bloom
Mental Health Clinic
United States Air Force Hospital Cannon (TAC)
Cannon Air Force Base, New Mexico 88101

Abstract

This paper presents two features of clinical psychology which may serve as a model for psychological operations (PSYOP) development. (1) A functional taxonomy of goal, focus, and level can order PSYOP data. (2) Kiesler's grid model can evaluate PSYOP data. Military lore, intervention research, and federally funded behavioral science projects are data sources which can be so ordered and evaluated. Although value differences may have limited contact between clinical psychology and PSYOP, rapprochement could provide a sound base for PSYOP as one of our four major weapon systems.

This paper will outline how psychological operations (PSYOP)--the military utilization of psychological phenomena--should be developed as one of the four major weapon systems (Gleason, 1971). It will also describe how two goals of clinical psychology can serve as signposts for PSYOP development.

Why even consider the healing entreprise of clinical psychology in the context of PSYOP? After all, clinical psychology primarily seeks to increase the well-being of troubled people through assessment, intervention and evaluation (Krasner, 1971; Rybstein-Blinchik, 1979). Yet there are sufficient commonalities to provoke curiosity as to why the two specialties have not enjoyed cross-fertilization, or even significant contact.

Both have been deeply associated with the military. PSYOP, of course, is by definition a military activity with a history as long as warfare itself (Breasted, 1944; Kitto, 1957). On the other hand, clinical psychology claims its beginnings in the late nineteenth century (Cattell, 1890; Galton, 1883; Witmer, 1907), but its most vigorous periods of development involved the creation and implementation of assessment instruments for the military in Worl Wars I and II (Watson, 1949; Yerkes, 1921).

A second link between clinical psychology and PSYOP is their deep involvement with the assessment, intervention, and evaluation of human functioning. Here human functioning subsumes both the therapeutic inclinations of clinical psychology and the military perspectives of PSYOP.

This link is accompanied by two distinct sets of values. Values that influence the nature of assessment. Values that can be the reason, cause, means, and end of intervention. Values that permeate the essence of evaluation (Kuhn, 1962; Natsoulas, 1978).

A major cause of the limited association between clinical psychology and PSYOP probably involves their respective values of intervention. Clinical psychology is imbued with the values of science and of art (Dimond et al, 1978; Haley, 1973; Rogers, 1942). Many clinicians state their independence from or the irrelevance of geopolitical and sociopolitical issues (Jacoby, 1975; The politicization of an academician, 1977)--either explicitly (Referendum backs council, 1979) or implicitly (May, 1968).

On the other hand PSYOP is imbued with political realities and pragmatics divorced from the luxurious notion of constantly acting for the Individual Good (Kitson, 1971). A Hegelian utilitarian philosophy is salient here. The PSYOP technician as clinician would be merely a rehabilitator, not also a purveyor of hope (Frank, 1961).

The results of limited association between PSYOP and clinical psychology has resulted in outright blunders that may not have occurred with basic contributions from the latter (Watson, 1978). Two goals of clinical psychology in particular could be most useful to PSYOP development and implementation. One involves the construction of a functional taxonomy. The other seeks to create suitable modes of evaluation.

Clinical psychology's functional taxonomy classifies intervention techniques according to the goal, the focus, and the level.

There are two main goals of intervention in clinical psychology's functional taxonomy. One involves intra- and extra-individual change. The other fosters maintenance of a current intra- or extra-individual status.

There are two main foci in the functional taxonomy--the individual and varying systems of individuals. These systems may involve a dyad, group, and formal or informal organizations.

Finally there are four levels of intervention. The biological level may involve electroconvulsive shock. The intrapsychic level can be illustrated by cognitive-behavioral approaches. The interpersonal level can include role playing. Lastly the social level may comprise direct intervention in the human ecology.

Of course, to decide whether change or status maintenance of a given focus at a given level is needed--or has been appropriately induced--, assessment must occur. Assessment devices span levels similar to techniques of intervention. Biochemical assays (the biological level), projective tests (the intrapsychic level), sociometric scales (the interpersonal level) and national surveys (the social level) represent some of these assessment devices. As with intervention levels, most assessment levels are interdependent.

If applied to PSYOP, clinical psychology's funxtional taxonomy could aid in engendering appropriate goals, foci, and levels of intervention for the PSYOP consumer.

For example with attitude change (an intra-individual goal) and a country's populace as the focus of intervention (a system focus), the social level of intervention--perhaps mass media communication--may be most appropriate. The assessment device of the survey (the social level) may indicate the current attitude of the populace and, indeed, may initially detect the necessity for attitude change by depicting a significant gap between the populace's contemporary attitude and an ideal compatible with United States politico-military policy.

From within the general class of mass media techniques, the promulgation of attitudes gradually approximating the ideal could be effective. This could be handled by a highly respected member of the given populace. Such an approach is analagous to successive approximation experiments in general psychology involving subjects' perseveration of color labelling in response to subtle wavelength shifts in nonambient light. Just as subjects maintain a "green" response as wavelengths gradually diverge from the 510-

530 millimicron range, citizens of the populace could maintain identification with the promulgator even as the communications diverged from the initial attitude. Here the PSYOP planner would choose the highly respected member as promulgator in adherence to social psychological data indicating increasing persuasion facilitation with increasing status of the promulgator (Hovland & Weiss, 1951). The PSYOP planner would also be aware that the perseveration process of identification leading to persuasion facilitation has limits, much as the perseveration of color labelling (Jacobs, 1976).

In considering this simple example, the discerning reader will have noted the following difficulties. First is the threat of using inappropriate increments of attitude. If the initial increment is too large, unforseen or even unwanted effects may occur (Hovland et al, 1957). On the other hand, it may be just the vast disparity between the promulgator's respected status and the unpopularity of the ideal attitude which engenders significant attitude change in the desired direction (Greenwald & Ronis, 1978). Furthermore, a promulgator with low status may easily change a populace's attitude, if the ideal is dramatically opposed to that embraced by the promulgator (Walster et al, 1966).

A second difficulty involves excessive attention on the ultimate target of intervention and the exclusion or insufficient emphasis of the human ecology. For example a PSYOP consumer may select a goal of eliciting valuable data from a noncompliant individual (an intra-individual change goal, an individual focus). The technique of intervention could be the introduction of persuasive information in an anxiety-provoking environment (an interpersonal level)--as anxiety can be a positive correlate of learning and acceptance (Aronson & Carlsmith, 1963). The assessment device of heart rate monitoring (a biological level) could be an indicator of sufficient anxiety, as increasing heart rate and anxiety are often positively correlated (Cattell, 1963). As heart rate would increase, the consumer would predict anxiety to increase, the persuasive information to be learned and incorporated, and the individual to reveal the valuable data. Not quite. Increasing heart rate can also improve cognitive functioning--facilitating thought and increasing resistance to persuasion (Cacioppo, 1979). An insufficient emphasis on the interrelationships of all relevant variables in the target of intervention's human ecology may lead to the very opposite of the desired goal.

A third difficulty is the implicit linear relationship between goal, focus, and level. In contrast, there have been many nonlinear models created to predict the effects of mass communication and related persuasion attempts. These are often Rasch models which presume to take into account individual differences and varying exposure to communiques replete with Poisson distributions (Kempf & Repp, 1977). The example in question, however, is more a schematic illustration of possibilities for clinical psychology-PSYOP linkages than a definitive portrayal of the "state of the art".

All these difficulties--as well as the usual issues of cross-cultural generality (Klineberg et al, 1979)--are indicative of problems of efficacy. Clinical psychology's history and "state of the art" in seeking suitable modes of evaluating intervention techniques (Frank, 1979) is the second

exemplar for PSYOP development.

An intervention technique can be judged by the percentage of applications yielding the desired effect (Landis, 1937). For example, technique A would be superior to technique B if the former had a greater ratio of successful to unsuccessful applications.

Intervention techniques may also be compared for degrees of change (Miles et al, 1951). Here technique A might engender a lesser number of successful applications, but each application would be more proximal to the desired effect than technique B. Or technique B might even include applications engendering the opposite of the desired effect.

A further refinement notes the reactivity of an intervention technique with an assessment device, evaluation mode, or nonspecific factors (Friess & Nelson, 1942). In these cases statistical comparisons with some base rate in the absence of intervention and with varying levels of assessment should occur.

Clinical psychology's most recent improvement in measuring efficacy involves discrediting the "uniformity hypothesis" (Kiesler, 1966). This hypothesis suggests that an intervention technique can be similarly applied by all agents of intervention; that all agents of intervention can function with similar biological, intrapsychic, interpersonal, and social characteristics regardless of the technique to be applied; that all foci of intervention with identical labels, e.g., schizophrenics or communists, would, indeed, be equivalent within that label.

The "uniformity hypothesis" has been rejected by those interested in the outcome and process of intervention techniques (Bergin & Garfield, 1978). In its place a "grid model" has been developed. This "grid model" suggests that for different combinations of goals and foci, different levels (and agents) of intervention would be most effective (Kiesler, 1969)

There are several sources of data which PSYOP can apply in the development of its own "grid model" and towards the construction of its own functional taxonomy. First, the anecdotes and lore of military history can be coded according to the goal, focus, and level of intervention. Anecdote and lore also have heuristic value--generating new ideas, theories, and experimental possibilities (Kohler, 1947).

A second source of data is the vast intervention literature in the behavioral sciences, e.g., social, personality, and biological, exploring potent and interacting variables. The effects of high ambient temperatures on the probability of collective violence (Carlsmith & Anderson, 1979), the personality correlates of susceptibility to biasing information (McFall & Schenkein, 1970), and the psychosocial dimensions of assasination (Westmeyer, 1973) are some research pursuits with relevance to PSYOP.

A third source of data could be generated by governmental mandate. This would involve relevant information retrieval from federally funded research in the behavioral sciences. Both civilian, e.g., the National Science Foundation, and military, e.g., the Defense Advanced Research Projects Agency, locations would be subject to the coding of information for PSYOP use--by goal, focus, and level. A larger portion of military research budgets including the Office of Naval Research, the Army Research Institute for the Behavioral and Social Sciences, and the Air Force Office of Scientific Research could also be allocated for PSYOP-relevant explora-

tions.

Creation of a functional taxonomy through suitable modes of evaluation should be the ultimate objective of any applied science. Until a coherent research and development policy is implemented for PSYOP, this objective will not be met. Given the significant cost-effectiveness of PSYOP (Walker 1977) and its demonstrated enhancement of other major weapon systems (Lansdale, 1972), the failure to meet this objective spells misfortune. Misfortune for United States military might in an era of imposing operational requirements and geopolitical crisis.

References

Aronson, E., & Carlsmith, J.M. Effect of the severity of threat on the devaluation of forbidden behavior. Journal of Abnormal and Social Psychology 1963, 66, 584-588.

Bergin, A.E., & Garfield, S.L. (Eds.). Handbook of psychotherapy and behavior change (2nd ed.). New York: Wiley, 1978.

Breasted, J.H. Ancient times: A history of the early world (2nd ed.). Chicago: Ginn & Co., 1944.

Cacioppo, J.T. Effects of exogenous changes in heart rate on facilitation of thought and resistance to persuasion. Journal of Personality and Social Psychology, 1979, 37, 489-498.

Carlsmith, J.M., & Anderson, C.A. Ambient temperature and the occurrence of collective violence: A new analysis. Journal of Personality and Social Psychology, 1979, 37, 337-344.

Cattell, J.M. Mental tests and measurements. Mind, 1890, 15, 373-381.

Cattell, R.B. The nature and measurement of anxiety. Scientific American, 1963, 212, 82-91.

Dimond, R.E., Havens, R.A., & Jones, A.C. A conceptual framework for the practice of prescriptive eclecticism in psychotherapy. American Psychologist, 1978, 33, 239-248.

Frank, J.D. Persuasion and healing. Baltimore: The Johns Hopkins Press, 1961.

Frank, J.D. The present status of outcome studies. Journal of Consulting and Clinical Psychology, 1979, 47, 310-316.

Friess, C., & Nelson, M.J. Psychoneurotics five years later. American Journal of Mental Science, 1942, 203, 539-558.

Galton, F. Inquiries into human faculty and its development. London: J.M. Dent, 1883.

Gleason, R.L. Psychological operations and air power: Its hits and misses. Air University Review, 1971, 22, 34.

Greenwald, A.G., & Ronis, D.L. Twenty years of cognitive dissonance: Case study of the evaluation of a theory. Psychological Review, 1978, 85, 53-57.

Haley, J. Uncommon therapy: The psychiatric techniques of Milton H. Erikson, M.D. New York: W.W. Norton & Co., Inc., 1973.

Hovland, C.I., Harvey, O.J., & Sherif, M. Assimilation and contrast effects in reaction to communication and attitude change. Journal of Abnormal and Social Psychology, 1957, 55, 244-252.

Hovland, C.I., & Weiss, W. The influence of source credibility on communi-

cation effectiveness. Public Opinion Quarterly, 1951, 15, 636-650.
Jacobs, G.H. Color vision. Annual Review of Psychology, 1976, 27.
Jacoby, R. Social amnesia: A critique of contemporary psychology from Adler to Laing. Boston: Beacon Press, 1975.
Kemuf, W.F., & Repp, B.H. Mathematical models for social psychology. Chichester, England: Wiley, 1977.
Kiesler, D.J. Some myths of psychotherapy research and the search for a paradigm. Psychological Bulletin, 1966, 65, 110-136.
Kiesler, D.J. A grid model for theory and research in the psychotherapies. In L.D. Eron & R. Callahan (Eds.), The relationship of theory to practice in psychotherapy. Chicago: Aldine Publishing Co., 1969.
Kitson, F. Low intensity operations: Subversion, insurgency, and peacekeeping. London: Faber & Faber, 1971.
Kitto, H.D.F. The Greeks. Baltimore: Penguin Books, 1957, 136-152.
Klineberg, O., Zavalloni, M., Louis-Guerin, C., & Ben Brika, J. Students, values, and politics: A cross-cultural comparison. New York: The Free Press, 1979.
Kohler, W. Gestalt psychology. New York: Liveright, 1947.
Krasner, L. The operant approach in behavior therapy. In A.E. Bergin & S. L. Garfield (Eds.), Handbook of psychotherapy and behavior change: An empirical analysis. New York: John Wiley & Sons, Inc., 1971.
Kuhn, T.S. The structure of scientific revolutions. Chicago: The University of Chicago Press, 1962.
Landis, C. A statistical evaluation of psychotherapeutic methods. In L.E. Hinsie (Ed.), Concepts and problems of psychotherapy. New York: Columbia University Press, 1937.
Lansdale, E.G. In the midst of wars. New York: Harper & Row, 1972.
May, P.R.A. Treatment of scizophrenia. New York: Science House, 1968.
McFall, R.M., & Schenkein, D. Experimenter expectancy effects, need for achievement and field dependence. Journal of Experimental Research in Personality, 1970, 122-128.
Miles, H.H.W., Barabee, E.L., & Finesinger, J.E. The problem of evaluation of psychotherapy with a follow-up study of 62 cases of anxiety neurosis. Journal of Nervous and Mental Disease, 1951, 114, 359-365.
Natsoulas, T. Residual subjectivity. American Psychologist, 1978, 33, 269-283.
Referendum backs council action on ERA. APA Monitor, May 1979, p. 1.
Rogers, C.R. Counseling and psychotherapy: Newer concepts in practice. Boston, Mass.: Houghton Mifflin, 1942.
Rybstein-Blinchik, E. Effects of different cognitive strategies on chronic pain expereience. Journal of Behavioral Medicine, 1979, 2, 93-102.
The politicization of an academician. Advance, November 1977, pp. 4-5.
Walker, F.W. PSYOP is a nasty term--TooBad. Air University Review, 1971, 28, 71-76.
Walster, E., Aronson, E., & Abrahams, D. On increasing the persuasiveness of a low-prestige communicator. Journal of Experimental Social Psychology, 1966, 325-342.
Watson, P. War on the mind: The military uses and abuses of psychology. New York: Basic Books, Inc., 1978.
Watson, R.I. The professional status of the clinical psychologist. In R.

I. Watson (Ed.), Readings in the clinical method in psychology. New York: Harper, 1949.
Westermeyer, J. Assasination in Laos: Its psychosocial dimensions. Archives of General Psychiatry, 1973, 74C-743.
Witmer, L. Clinical psychology. Psychological Clinics, 1907, 1, 1-19.
Yerkes, R.M. (Ed.). Psychological examining in the United States Army. (Memoirs of the National Academy of Sciences, Vol. 15). Washington, D.C.: U.S. Government Printing Office, 1921.

Footnotes

1 The author is indebted to Colonel William H. Berrick, USAF, Director of Base Medical Services, Cannon AFB, NM, and to the staff and students of Squadron Officers School, Class 79-B (especially Section D-45), Maxwell AFB, ALA, for fostering an environment conducive to creativity.

EXAMPLES OF THE BROADENED ROLE OF PSYCHOLOGY IN HEALTH MAINTENANCE OF NEUROENDOCRINE PATIENTS

Lawrence F. Wilson, Ph.D.
USAF Medical Center
Wright-Patterson AFB, Ohio 45433

Abstract

In medical facilities, psychologists are no longer asked to practice solely as psychotherapists. They are able to and often do assume multiple roles in health organizations. This paper describes the activities of a USAF psychologist (the author) in a major medical center in the broadened role of Health Psychology. The incidents are portrayed by roles performed in the health maintenance of neuroendocrine patients who are followed by internists and endocrinologists. Three cases ((1) Status Post, Craniopharyngioma, (2) Graves disease, and (3) Addison's disease) are presented which reflect the services and roles (clinical consultant, researcher, clinician) offered by the health psychologist in several internal medicine speciality clinics. The basis for the roles is an extension of clinical neuropsychology and its methods. The author contends that there is a challenge and diversity of function for psychologists who intend to respond to the need for concentration on the delivery of high quality health care.

The involvement by the newly formed Division of Health Care Psychology (Division 38) of the American Psychological Association demonstrates the growing interest of psychologists to utilize their academic and clinical training and skills in health care management and health care settings. The greater heterogeneity among those seeking treatment than exists at mental health settings has led to descriptions and development of programs and roles among psychologists emphasizing cost-effectiveness, prevention, and brevity of treatment (Weisenberg, 1970; Schofield, 1979). Contributions to the literature and books in the field concerned with mental health services in health maintenance settings were discussed in a Special Issue of Professional Psychology (August, 1979).

Health Care Psychology, as a discipline, emphasizes a strategy for operational and theorectical values of research and exchange of information in health transactions. Role definition emphasizes the psychological aspects of the health care process - the cognitive, social psychological and personality issues which determine or influence outcomes in health transactions. Within the medical environment, roles are acted upon through offering consultation regarding the psychological problems of individual patients, teaching medical students, and developing research collaborations with persons in health care settings about questions of concern in delivery of high quality health management, and contributing to psychological theory.

The clinical psychologist, because of his training and exposure, can concentrate on the delivery of health care and particularly those aspects of health care that involve physicians. In his role as health care psychologist, the clinical psychologist will need to rely on his strength in the content areas of physiological psychology, clinical neuropsychology,

PRECEDING PAGE BLANK-NOT FILMED

clinical research design and methodology, personality theory, theorectical models of behavior and learning, and the broad range of rehabilitation and psychotherapy. Criteria for this concentration are (a) intention to spend professional life in interdisciplinary settings, (b) intention to become a research-participant who consults and helps to solve operational problems on the basis of data and data-based theories, (c) desire to teach students in other health professions and (d) desire to become actively involved in the social and biomedical aspects of the medical environment. Role diversity available to the clinical psychologist who chooses this concentration may include: Administration/liaison, clinical consultant, teacher, researcher, clinician, and organization consultant.

Case examples of clinical psychologists who have focused in the broadened role of health psychologist and concentration on delivery of health care should be of interest.

Method

Three roles are discussed which demonstrate the activities of a USAF psychologist, the writer, as a health psychologist in collaboration with internal medicine specialty clinics and the health maintenance of neuroendocrine patients. The roles are: Clinical consultant, researcher and clinician. Role descriptions and case examples are presented to elucidate the health psychology relationship with medical health care practice of physicians.

Results

Clinical Consultant. Provides psychological services as resources for health care providers who are having difficulty with particular patient management, because of significant clinical assessment of behavioral or emotional problems. Assists providers in counseling approaches to patients with serious illness as well as comprehensive approaches to family involvement approaches as it affects health care delivery.

Case Example. Mr CT is a 47 year old male with status post resection of craniopharyngioma. His physician, an internist, stated that in the health management of this individual, it was observed that the patient was having difficulty accepting his condition. Also, the patient and his spouse exhibited marked depressive symptoms. The request to the health psychologist was to assist in evaluation of both the patient and his wife, to determine baseline neuropsychiatric functioning and assist in prescribing an approach to use in resolving the conflicts surrounding this health management situation. A neuropsychological evaluation of the patient and an interview and psychological testing of his wife was provided for baseline functioning data. The psychologist, together with the internist provided specific information to the patient and his wife about psychological aspects of psychosomatic problems with regard to treatment, a brief description in graphic and dydatic terms of the physiology and process of the illness. Specific recommendations in terms of rehabilitation utilizing good functioning areas of the brain were provided. Referral to a rehabilitation agency was made. Explanation of the interactive effects of the management problem on the physician and patient was offered, with all participants engaged in the process. Throughout the consultation the internist remained the primary health care provider.

Researcher. Conducts research related to clinical practice and evaluates the efficacy of medical treatment methods as well as impact of

psychological and mental health services on medical utilization. Investigates and determines relevant variables before, during and after medical interventions.

Case Example. Mrs AS is a 55 year old female widow who is status post Graves disease presently in euthyroid status. Her treatment was by radioactive iodine ablation therapy, inderal medication for tachycardia and synthyroid replacement. She was seen in the thyroid specialty clinic for follow-up. The endocrinologist noted that while she was biochemically and clinically determined euthyroid, there appeared to be significant psychological factors which influenced the management of this patient. The endocrinologist called upon the health psychologist to explore the possibility of determining a profile of relevant variables to better understand the on-going process of Graves disease. He was aware of past studies, but found the data in his literature to be of little practical use in his clinic. The health psychologist, together with endocrinology and nuclear medicine, designed an interdisciplinary approach to determine brain-thyroid-behavioral relationships which are associated with Graves disease, thyrotoxicosis and treatment. Results would be considered for publication in a medical journal of interest to endocrinologists and behaviorists.

Clinician. Provides psychotherapy to individuals, families, couples and groups with health plan members who are referred by the primary care provider. Focuses on brief psychotherapies with an emphasis on cognitive-behavioral models. Assists in determination of appropriate referrals.

Case Example. Mr RW is a 64 year old male with surgically induced Addison's disease who exhibits marked apathetic depression, low frustration tolerance and agitation. The endocrinologist, before contacting the health psychologist, had been involved in intense and supportive counseling to the patient. He and the patient had discussed the symptomatology of Addison's disease. The health provider felt impotent in light of the extent of the behavioral dysfunction and apathy shown by the patient. His referral to the health psychologist was for psychotherapy, but the ednocrinologist would remain in the role of primary health care provider. A neuropsychological evaluation was accomplished and results were provided to the endocrinologist prior to treatment by the psychologist. A cognitive-behavioral approach with the target being depression was used by the psychologist over 10 sessions. The patient responded well to psychotherapy and the sessions terminated. Follow-up projective psychological testing and a report was accomplished. The endocrinologist continued as primary health care provider.

Conclusions

Psychologists are no longer asked to practice solely as psychotherapists. They are able to and often do assume multiple roles in health management organizations with a focus on health care delivery. The activities described in this paper demonstrate a multifaceted interdisciplinary role of a USAF psychologist in a major Medical Center. The health psychologist provided a service which concentrated directly on the needs of physicians who provide primary health care delivery.

As psychologists move into expanded roles in medical environments, challenges are realized both within and among fellow professional organizations and other professions (Rosen and Wiens, 1979). The opportunity

to conduct research related to clinical practice and evaluate the efficacy of treatment methods, either medical or psychological is rewarding and should provide the basis for contributions to psychological theory.

References

1. Rosen, J., & Wiens, A. Changes in medical problems and use of medical services following psychological intervention. American Psychologist, 1979, 34, No. 5, 420-431.
2. Schofield, W. The clinical psychologist as a health professional. In G.C. Stone, F. Cohen, & N.E. Adler (Eds.), Health Psychology. San Francisco: Jossey-Bass, 1979.
3. Special Issue: Psychologists in health care settings. Professional Psychology, August, 1979, 10, No. 4.
4. Weisenberg, M. The role of psychology in the delivery of health services. American Psychologist, 1970, 25, 472.

Selected Factors Associated with Success in a Behavioral Weight Loss Program

Darrel Parisher
Mental Health Clinic
USAF Hospital
Holloman AFB, New Mexico 88330

Abstract

This study investigated whether a relationship existed between successful completion of a behavioral weight loss program and selected personality, behavioral, and demographic variables. A secondary goal of the study was to determine whether a relationship existed between this same set of variables and an individual's percentage of body weight lost. Thirty-one university subjects completed a behaviorally oriented weight control program which lasted six weeks. Data analysis using a correlational approach revealed few significant correlations, but ancillary findings showed a number of significant factors. In discussing the results it was concluded that although trying to ascertain a list of characteristics of successful weight losers is difficult, this type of research is indeed important in designing future programs.

Losing weight seems to be a preoccupation of the American people. Estimates are that 25-45% of the adult American population over age 30 is more than 20% overweight (Bray, 1976). Our society seems to stigmatize the overweight person as being handicapped, and overweight has been viewed in religion as a sin, in medicine as a disease, and aesthetically as ugliness (Allon, 1973). Diet books are among best sellers (Kimbrell, 1975) and the diet industry is a $10,000,000,000 business (Allon, 1973). The problem of being overweight is complicated by the fact that 95% of all cases of obesity have no apparent medical cause (Rodin, 1978). The old excuses of having glandular problems, fat genes, or a large bone structure have very little support in the data. Given the fact of a large number of overweight people and a primarily psychological etiology, it is logical that many approaches to helping people lose weight are psychological in nature and based on the literature of psychology.

Leon (1976) reviewed the literature on the treatment of obesity and reported that behavioral methods, individual psychotherapy, hypnosis, diet approaches, drugs, and surgery have all been used as treatment approaches. She concluded

that with the exception of surgical intervention, behavior modification techniques generally have proven to be the most effective treatment procedure. The use and success of behavioral techniques have been reported by Stunkard and Mahoney (1976). One method which has shown considerable success is behavioral techniques combined with instruction about diet and exercise (Stuart & Davis, 1972). Although behavioral methods have been successful, they do not seem to work on all subjects. Jeffrey (1974b) stated that since behavioral methods do seem to be successful, much could be gained by studying what factors are involved in successful and unsuccessful attempts at weight control.

Several authors have noted that there seems to be an emotional component involved when someone is overweight (Kiell, 1973; Mayer, 1968; Schachter, 1971a; Tyson & Tyson, 1974). Leon and Roth (1977) reviewed a great deal of the literature concerning psychological factors associated with obesity and concluded that:

> The evidence strongly suggests that there are very few personality characteristics that obese persons share that can be considered causative in the development of obesity. Neither orality, externality, nor depression universally characterize the obese individual. Although some obese persons have been shown to be sensitive to external, environmental cues, they have also demonstrated a greater degree of emotionality. The search for a unitary explanation of obesity does not, at present, appear to be fruitful avenue of exploration, and the evidence strongly suggests that obesity is not a unitary syndrome. (p. 136).

Although the authors provide little encouragement to researchers in this area, much effort has been exerted in trying to determine effectiveness.

In general, a conclusion suggested from reviewing the weight loss literature is that although behavioral methods have been effective, there is little known about what factors contribute to successful attempts at weight reduction. There has been limited research into this area, and what has been done would best be described as inconclusive. The present study attempted to examine the possible relationship between an individual's successful completion of a behaviorally oriented weight loss group and selected personality, behavioral, and demographic variables. A secondary aspect of the study was to determine if an increase in the amount of weight lost by an individual had any relationship to the same set of variables.

Method

Subjects. Forty-two students at a Southwestern university took part in this research. Of the 42 subjects, 41

were female and the lone male completed the program but was not used in the final data analysis. The mean age of the participants was 20.4 years and their mean weight was 150.6 pounds (68.31 kgs.) Subjects were volunteers who responded to an advertisement asking for individuals who were interested in losing weight.

Instrumentation. The trait scale of the State-Trait Anxiety Inventory (STAI) (Spielberger, Gorsuch, & Lushene 1970) was used to determine the amount of trait anxiety present for each participant. The Profile of Mood States (POMS) (McNair, Lorr, & Droppleman, 1971) measured participants state disposition on several factors. These factors are depression, tension, anger, vigor, fatigue, and confusion, and several of these have been considered important by researchers. The K-scale of the Minnesota Multiphasic Personality Inventory (MMPI) (Dahlstrom & Welsh, 1943) was used primarily to measure self-esteem. Locus of control, as measured by the Attitude Statement Survey (Levenson, 1972, 1974) has been presented by various researchers as an integral part of determining weight loss resolution. A Weight Loss Questionnaire was constructed by the investigator and used to gather behavioral and demographic data about each subject. The questionnaire was designed by reviewing the weight loss literature and also reviewing other questionnaires used for behavioral assessment. Several considerations were involved in obtaining material for the questionnaire. Motivation for weight loss, history of weight problem, family overweight pattern, prior attempts at weight loss, and general demographic characteristics were primary concerns.

Procedure. Three treatment groups or sections were formed and all used the same treatment approach. The investigator led all three groups and was assisted by a different co-therapist for each group. Groups were mainly didactic in orientation and the leaders served to disseminate information and to instruct. Group one had 12 members and groups two and three each had 15 members. Each group lasted one and one-half hours per week and met for six weeks. A modification of the plan devised by Stuart and Davis (1972) was the treatment program used. Modification consisted of adding one session of relaxation training to the procedure. The program is based on the behavioral procedures of self-management and stimulus control with the addition of instruction in diet and exercise.

Results. Thirty-two subjects met the program criteria of attendance at all or all but one group session and finished the program. Twenty-one participants lost weight during the program. A completion rate of 76% was therefore realized. The weight of completers ranged from 114 to 198 pounds (51.71 to 89.81 kgs), and the mean was 147.47 pounds (66.89 kgs) at the end of the treatment period. In terms of pounds (kgs) of weight lost, the mean was 4.38 pounds (1.99 kgs). In order

to assess the major aims of the study, the percentage of body weight lost for each participant who lost weight was computed. The percentages ranged from 1% to 7%, with the median of this distribution being 3%. The median served as the cut-off between the successful and unsuccessful participants in the weight loss program.

In the final data analysis, no significant correlations were found except for the confusion scale of the POMS. For both successful completers and those who lost more weight, there tended to be less confusion experienced. No significant correlations on any of the behavioral or demographic variables was found. However, in a study of this type there are relationships between variables which though not part of the major thrust of the study are worth mentioning.

Much research in the weight loss literature has centered on determining the characteristics of people who are overweight. This investigation found several significant correlations which relate to this issue. A significant correlation was found between an individual's weight and her perception that her weight problem was severe, the years she has been overweight, and the number of attempts to lose weight. Results of this type suggest that being overweight is in many cases a long term proposition and that although the individual perceives the problem as serious, attempts at solution are often ineffective. On the average in the present study each participant had attempted to lose weight over five times previously.

Prior attempts at losing weight constituted another important factor in the present study. Severity of the weight loss problem was significantly related to the number of attempts a participant had made to lose weight. Other significant correlations with prior attempts were number of years overweight, the mother's being overweight, and agreement that being overweight has been a persistant problem.

A third set of significant correlations was found between the severity of the problem and a number of variables. Two personality measures were significantly related to the severity of the weight loss problem. The STAI was related to severity as well as the Depression scale of the POMS. These results suggest that there is a degree of anxiety and depression associated with an individual's feeling that his or her weight problem is severe. Other variables which were significantly related to severity were years overweight and someone demanding that they lose weight. What these data seem to imply is that the more years a person is overweight the greater the problem, also the greater amount of pressure they receive from others to lose weight.

The major correlations seem to be found under three variables: weight of the person, their perception of the severity of the problem, and the number of prior attempts.

In summary, these findings tend to suggest that the factors of a person's weight, prior attempts, and problem severity could be important factors in attempting to get a better understanding of the overweight person with possible implications for devising more effective treatment strategies.

Discussion. The results of the present study suggest that the selected personality, behavioral, and demographic variables had no significant relationship to either success in weight loss or amount of weight lost. An exception to this was the Confusion scale of the POMS which was significantly correlated to both factors, but as a whole the variables used were not useful in differentiating between groups. Findings tend to agree with Leon and Roth's (1977) conclusion that studies which attempt to find relationships between weight loss and other variables have largely had mixed results. Although many studies have found significant results, an equal number have reported unsuccessful resolution of the question. While significant findings were few in answering the major hypotheses of the investigation, several correlations were determined which provide some speculation about individuals who are overweight. In general, results suggest that overweight individuals perceive their weight as a pressing problem and have made numerous attempts to try to resolve the problem. These results, though tentative in nature, have some support from other research (Murray, 1975; Seaton & Rose, 1965; Waite, 1974; Wunderlich & Johnson, 1973).

The failure of this study to determine many significant relationships could be explained in several ways. Obviously the first of these is that research into individual differences involved in any procedure is extremely difficult to determine. Given the fact that this kind of research is filled with methodological and theoretical pitfalls, there still seems to exist a need for this type of investigation. One possible explanation for the equivocal findings of research into this question is that exploratory studies such as thing should be used to help build a data base for further research.

One methodological problem of all weight loss research is that replication is so difficult, usually because the type of sample differs so much from study to study. When viewing even a limited part of the weight loss literature, e.g. factors related to successful or unsuccessful treatment, sample descriptions can range from university women to weight club participants to adolescent girls. Results become complicated and hard to interpret since groups are so divergent, therefore variables which work with one group may not work with another. Even with the essentially negative findings of this study, continued investigation seems to be indicated. Additional information about who best can be served by particular therapeutic techniques seems warranted.

Inferences About Pain Patients

Rex A. Frank
USAF Hospital, Homestead Air Force Base, Florida

Abstract

The stereotypes held by physicians, nurses and the general public about chronic and acute pain patients and general medical patients were studied. After reading a stimulus paragraph and symptom list which described one of these types of patients, each subject completed the Adjective Checklist, checking terms which might best describe the "patient". Inferences about chronic pain patients showed significantly greater abrasiveness and less positive functioning than other patients across subject conditions. Results were discussed in terms of implications for the health professional-pain patient interaction and suggestions for treatment of pain patients and their families are made.

With the development of behavioral treatment methods for chronic pain, new emphasis has been placed on non-physiological factors which influence a patient's perception of pain, and the impact of pain on his daily behavior. Authors such as Shealy (1976) and Sternbach (1974) have identified the interaction between the health professional and the pain patient as one important factor in treatment. Shealy (1976, p. 1) states that most physicians who encounter chronic pain patients are frustrated by both the patient and the failure of traditional treatment. Sternbach (1974, p. 77) contends that anyone familiar with transactional analysis will have no problem in recognizing and counteracting the moves of the patient in pain games.

Berne's (1964) work with transactional analysis reveals that all participants in a transaction must be accounted for. While there is substantial data regarding the attitudes and motivations of pain patients in this type of transaction, few have investigated the health professional's role in the pain game.

Engel (1959) emphasized the meaning which pain may acquire for the patient in childhood. No literature was found which relates the meaning which pain may acquire in the physician's childhood to his treatment of his patients. Lenburg, Glass, and Davitz (1970) demonstrated that the race, sex, and cultural background of the physician all influence the meanings which he attaches to pain and suffering. Ultimately such meaning influenced inferences about medical patients in this study. However, few questions about pain patients were answered by this work. Bond and Pilowsky (1966) demonstrated that the treatment of medical patients may be influenced by factors other than the patient's actual complaints and symptoms of pain. In another study, (Pilowsky, 1967) the patient's age, self concept, and sex all influenced staff treatment. Thus, health professionals appear to hold certain meanings of pain and tend to make predictable inferences about their patients which influcence treatment. The purpose of the present study is to investigate such inferences in the case of chronic pain patients.

PRECEDING PAGE BLANK-NOT FILMED

Method

Subjects

Sixty physicians, 60 nurses, and 60 members of the general public served as subjects. Physicians and nurses were selected from a large Veterans Administration/university medical center, while general public subjects were randomly selected from the phone directory in the same town.

Materials

Stimulus materials consisted of three separate paragraphs describing a 45 year old laborer's interests and activities, followed by a list of four physical and behavioral symptoms of illness. Stimulus paragraphs were identical. Symptom lists described a chronic pain syndrome, an acute pain syndrome, and a general medical problem. Chronic and acute pain symptom lists differed only on the basis of length of time in pain (seven years vs. three days). Symptoms of "moderate fever and nausea" were used in place of pain for the general medical patient stimulus. All behavioral symptoms were identical among stimuli.

Procedure

The three stimulus packages were randomly distributed within each observer group. After reading the stimulus, subjects checked adjectives on the Adjective Checklist (ACL) (Gough & Heilbrun, 1965), which they felt might describe the individual identified in the stimulus paragraph. Subjects also completed an attitude survey which is not examined in this study.

Analysis

The factor structure of the ACL was determined using principle components extraction and varimax rotation (Nie, Hull, Jenkins, Steinvrenner, & Bent, 1975). The mean of each factor score was calculated for each cell of the stimulus matrix. A two-way ANOVA (stimulus X observer) was performed for each factor, and the Newman Keuls test was applied to significant effects. A step-wise multiple discriminant analysis (Nie, et al., 1975) was performed to determine the accuracy of stimulus classification by observer groups. The accuracy of classification was tested with the Chi Square statistic.

Results

The first four factors accounted for 27% of the variance of the total analysis, and are similar to those found by Parker and Veldman (1969). Factors were labeled as follows: I) Positive Functioning, II) Abrasiveness-Anxiety, III) Passivity, and IV) Assertiveness.

Significant differences among stimuli (across all groups of observers) occur in both Positive Functioning ($p=.0009$) and Abrasiveness-Anxiety ($p=.0001$). Significant differences among observer groups were found only on Abrasiveness-Anxiety ($p=.0008$). Post hoc tests show chronics to evidence less Positive Functioning than general or acute patients ($p < .01$). Chronics also show greater Abrasiveness-Anxiety than the other patient groups ($p < .01$). The public attributed greater Abrasiveness-Anxiety to all patient groups than did physicians and nurses ($p < .01$). Figure 1 summarizes the post hoc tests on stimulus groups.

Discriminations are significant for stimulus ($p=.0001$), observers ($p=.001$), physicians ($p=.001$), and nurses ($p=.038$). Discriminations approached significance ($p=.076$) for the public sample. Physicians accurately classified 60% of chronics; nurses classified 45%, and the public 55%. Overall accuracy was 51.67%, 46.67% and 41.67% respectively.

Figure 1

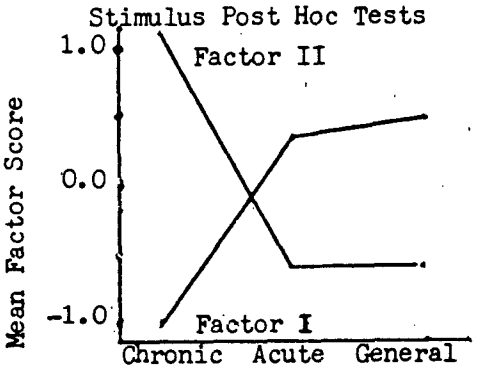
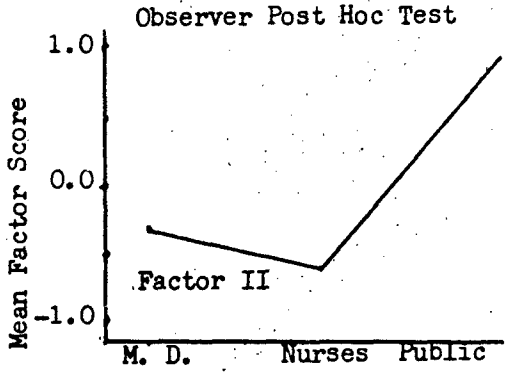


Figure 2



Review of the step-wise procedure for physicians reveals that discrimination is made among all three stimuli ($p < .05$). Factor I discriminates between chronic and general stimuli and acute and general stimuli. Factor II discriminates between chronic and acute and between chronic and general.

A similar review of the step-wise procedure for nurses indicates that they discriminate only between chronic and general stimuli to a significant degree ($p < .05$) Factor I provides the greatest discriminative power for physicians (weighting coefficient = .72401), while Factor II has greater power for nurses (weighting coefficient = .98102). Figure 2 summarizes observer post hoc tests.

Discussion

As expected, physicians and nurses ascribed high levels of neurotic personality traits to chronic pain patient stimuli, even when provided with length of time as the only distinguishing factor between chronic and acute stimuli. Surprisingly, however, the public sample also ascribed these traits. The ability of the public sample to discriminate chronic patients from others is shown by their 55% accuracy rate in the discriminant analysis. Though the overall classification function for the public is non-significant, this may be attributed to the extremely poor classification of the general medical patient stimuli (25%). The public actually exceeded nurses in their ability to discriminate chronic and acute stimuli.

Since the mean factor scores for general and acute stimuli are both closer to zero than are loadings for chronic stimuli on Factors I and II (Figure 1), it is concluded that physicians, nurses, and the public did not ascribe significant personality characteristics to the acute and general stimuli. However, the limitations imposed by only four personality factors extracted from the ACL do not allow for a strong conclusion.

It is apparent that physicians and nurses differ in their criteria for making discriminations among stimuli. Physicians based their distinctions primarily on the absence of Positive Functioning, while nurses based their distinctions on the presence of Abrasiveness-Anxiety. There is no evidence in this study to suggest the source of this difference. However, it is likely that nurses act as front-line troops in the management of pain patients, with greater direct contact. The physician has relatively little contact with the patient and greater prestige. The patient may be more

likely to emit negative behaviors in the presence of the nurse, leading to her basis for discriminating. When seeing the physician, the patient will probably be hesitant to express his negative characteristics since the physician ultimately controls his medication and treatment. However, the patient may be unable (or unmotivated) to show evidence of positive functioning before his physician, thus leading to the M.D.'s perception of its absence.

The evidence of greater predictive accuracy among physicians suggests greater consistency of stereotyping among those in this professional group. Again, no source of this consistency is evidenced in the present research, though we might hypothesize that it is a product of contact with other physicians and formal medical training. It is of note, however, that the public shows greater consistency than nurses in their classification of stimuli. In other data from this study, reported elsewhere (Frank, 1978), the public sample also showed greater discrimination of chronic patients than nurses based on their attitudes towards the "patient". In that analysis, the public's discrimination was significant ($p=.002$), and 60% of chronics were accurately classified on the basis of attitude. It is apparent that there is clear cut stereotyping of chronic pain patients by the public as well as health professionals.

This surprising result may be accounted for by the public's contact with pain patients. They may generalize from their experiences with individuals who have any type of chronic debilitating medical disorder that does not improve with time. They may also be frustrated in dealing with individuals whose major complaint in life does not improve.

This study demonstrates that physicians, nurses, and the public all have definite stereotypes about the chronic pain patient. It is tempting to interpret such inferences as a direct reflection of the actual personal characteristics demonstrated by chronic pain patients on such inventories as the MMPI. The work of Sternbach, Wolff, Murphy, and Akeson (1973_a and 1973_b) has shown that chronic and acute pain patients produce remarkably similar personality patterns on the MMPI. However, the present investigation shows that inferences about chronics are considerably different from those made about acutes. The findings also imply that intensive interaction between the patient and the health professional is not necessary for the development of negative stereotypes. Inferences developed irrespective of contact with the patient may account for a major portion of the frustration encountered in treating pain patients. If the patient is helped to understand that his report of long-term pain alone is likely to elicit negative responses from others, he may be more able to select behavior which minimizes such responses. The reduction of stereotypic responses from family members may also be of import. Though the patient improves with treatment, the family may not, and may lead the patient back into pain behavior upon his return home from treatment.

Sternbach points out that the pain patient-health professional interaction is characterized by the patient's attempts to confirm his identity as a suffering person. Even before the onset of pain the patient is likely to have developed his own inferences about people with pain problems. His increasing anxiety and depression over his pain may be partly a function of self-perception. The failure of operant treatment methods may be partly

attributed to absence of treatment for the patient's self-concept.

Stereotypes regarding pain patients develop long before the health practitioner's training or contact with the patient. Patients do not deserve treatment which is partially based on stereotypes which do not parallel their conditions empirically. Insight into preconceived attitudes and inferences may help to defuse the pain game from the health professional's perspective. Armed with such knowledge, the practitioner may be able to guard against undue influence of such stereotypes and attitudes on his interaction with the patient. Hopefully such awareness leads to more open interactions with patients and to better treatment.

References

Berne, E. Games people play. New York: Grove Press, 1964.

Bond, M.R., & Pilowsky, I. Subjective assessment of pain and its relationship to the administration of analgesics in patients with advanced cancer. Journal of Psychosomatic Research, 1966, 10, 203-208.

Engel, G.L. "Psychogenic" pain and the pain-prone patient. American Journal of Medicine, 1959, 6, 899-918.

Frank, R.A. Attitudes towards and inferences about pain patients: The observer's perspective. (Doctoral dissertation, University of Missouri, 1977). Dissertation Abstracts International, 1978, 38, 77-31721B. (University Microfilms No. 77-31721B).

Gough, H.G., & Heilbrun, A.B. Jr. The Adjective Check List manual. Palo Alto: Consulting Psychologist Press, 1965.

Lenburg, C.B., Glass, H.P., & Davitz, L.J. Inferences of physical pain and psychological distress (II). In relation to the stage of the patient's illness and occupation of the perceiver. Nursing Research, 1970, 19, 392-398.

Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K., & Bent, D.H. Statistical package for the social sciences (2nd ed.). New York: McGraw Hill, 1970.

Parker, G.V.C., & Veldman, D.J. Item factor structure of the Adjective Check List. Educational and Psychological Measurement, 1969, 29, 605-613.

Pilowsky, I. Dimensions of hypochondriasis. British Journal of Psychiatry, 1967, 113, 89-93.

Shealy, C.N. The pain game. Milbrae, Calif.: Celestial Arts, 1976.

Sternbach, R.A. Pain patients traits and treatment. New York: Academic Press, 1974.

Sternbach, R.A., Wolff, S.R., Murphy, R.W., & Akeson, W.H. Aspects of chronic low back pain. Psychosomatics, 1973_a, 14, 52-56.

Sternbach, R.A., Wolff, S.R., Murphy, R.W., & Akeson, W.H. Traits of pain patients: The low back "loser". Psychosomatics, 1973_b, 14, 226-229.

Unrecognized Factors and Assumptions Affecting Evaluations of Male and Female Officers

Lois DeFleur
Washington State University

Abstract

Two general issues concerning evaluations of male and female officers are discussed. The first is the need to recognize possible confounding effects of environmental factors such as the nature of group tasks and attitudes of work participants in differential evaluations of males and females. The second issue concerns the recognition of possible effects of sex-linked stereotypes and biases in the workplace. A number of examples of these are discussed as illustrations of potential pitfalls in performance evaluations.

All organizations, including the military, are concerned with assessing organizational as well as individual effectiveness and performance. Recently, there has been some discussion of assessment procedures in the military in light of a number of changes in these organizations. Janowitz (1971) and others have pointed to the changing nature of many military tasks and the social characteristics of members. For example, in recent years the demand for skills related primarily to leading troops has declined and there has been an increase in the need for technological sophistication as well as administrative and personnel management skills. At the same time, individuals from a wider range of backgrounds have been recruited and there are more minorities and women in the military than in the past (Binkin and Bach, 1977). In light of these trends it is particularly important that some of the issues and assumptions surrounding performance evaluations be examined. It may be that certain procedures will affect various military personnel in different ways and/or that certain criteria need not be as salient as they once were.

I would like to explore two general issues in the context of evaluations of male and female military participants. The first concerns the need to consider possible confounding effects of environmental factors in evaluations of individual performance. Several examples of such factors should suffice to illustrate this issue. An important environmental consideration which needs to be taken into account when evaluating performance is the basic nature of the task. Importance of task is well documented in the leadership literature (Stogdill, 1974) but I would like to discuss this issue in terms of male and female officers. For example, I think it is important to consider whether officers have been engaged in tasks which are stereotypically evaluated as appropriate or inappropriate for their sex and whether in fact such tasks actually have been highly sex-segregated in the past.

One of the consequences of stereotyping tasks or other activities is that people attribute various kinds of supplementary characteristics and traits to individuals engaged in these activities. This is particularly true when the tasks are not regarded as appropriate for these individuals. For example, in the area of deviant behavior, we find that a large number of auxiliary characteristics are attributed to women who engage in criminal

PRECEDING PAGE BLANK-NOT FILMED

behavior. Such women are regarded as aggressive, dominant, etc. indicating the imputation of a number of masculine traits as a result of their activity (Phillips and DeFleur, 1980).

In the study of group tasks and sex of leader, recent research has found that certain personal qualities were attributed to the leader and interpretations were made concerning the nature of his or her behavior depending on the relationship of task and sex (Calder, 1977).

A study conducted at West Point found that the nature of the task influenced group performance as well as leader evaluations (Rice, et al 1980). In this research, mixed-sex groups were asked to carry out stereotypically masculine or feminine tasks and results indicated that perceived masculinity of tasks resulted in differential demands on leaders which in turn were related to attributions concerning the leaders. This led researchers to conclude that the nature of the task and, more generally, the structure of the situation had strong effects on both group performance and individual attributions.

Another dimension of this issue is whether certain tasks have, in fact, been highly sex-linked and whether males and females have had similar background experiences and training. Such situations create initial advantages and disadvantages for each group and are likely to influence evaluations. For example, studies of males and females in pilot training have shown that the learning curve for women initially rises more slowly than for males. Such a situation could affect initial performance evaluations. This points to the need for evaluations with components tapping potential as well as actual performance.

A second dimension of the environment which is likely to influence individual evaluations is the nature of the attitudes of those in the immediate setting. In most work situations the relevant individuals would be supervisors, co-workers or followers. In terms of males and females in various roles in these settings, we find evidence to support the notion that sex-role attitudes of male supervisors and co-workers affect evaluations of females. In addition, leadership studies have found that followers' attitudes play a crucial role in attributions about leaders (Deaux, 1976). In the recent study at West Point, there was a strong negative bias in evaluations of female leaders by followers with traditional attitudes toward changing roles for women (Rice, et al., 1980). These attitudes were reflected in expectations about the leaders' aptitudes and behaviors and affected group performance.

The impact of attitudes and beliefs is also illustrated in a recent study of men and women in the Army Signal Corps (Rustel, 1980). Interviews were conducted with male and female members of the corps as well as their supervisors. One of the significant findings was that male supervisors evidenced very strong, essentially negative, orientations concerning females in their units. In addition, interviews indicated that a number of male co-workers saw women as nuisances within their work groups and felt females had inherent shortcomings which should disqualify them for such positions. The investigator concluded that such attitudes and evaluations contributed significantly to the role strain evidenced by young women in these units. These women soldiers were discouraged, dissatisfied with their job, felt they were not treated fairly and were up against impossible odds in performing their assignments. According to the author of the

study, "it is not surprising that women occupying such nontraditional jobs resolved these role conflicts by leaving the military. Both of these studies clearly indicate the necessity of being cognizant of various orientations within a work group and the possible impact on performance and evaluations.

The other general issue which I would like to discuss briefly is the need to examine evaluation criteria and procedures in terms of unrecognized assumptions and biases that may be operating. Everyone subscribes to a number of these, but often if such biases are made explicit or discussed it may help attenuate their influence or even eliminate them.

For example, I find it interesting that in discussions of work, family, dual careers and the like the focus is consistently in terms of women. This is easy to understand, given the structure of work organizations dominated by males and home settings occupied by women, but I think this implies that such issues affect primarily women which is not always the case. For example, I have heard a number of male military officers talk about problems of work and family and careers of spouses which affect their assignments. I think that males, as well as females, should be queried concerning these issues. Thus, I am simply pointing out that to assume that problems of integrating work and family are associated only with females is unwarranted.

Another example of unrecognized assumptions can be found in many sex-linked stereotypes surrounding work and performance. A typical assumption is that women are not committed to their work or job in the same way as males. Studies have found that the type and level of commitment varies within sexes and it is unwarranted to conclude that females always are less committed (Acker, 1978).

Another interesting assumption is that women typically lose time from work because of distinctive female-linked biological problems. In the American society, we have ready-made cultural categories for interpreting certain mild illnesses of women and these also have an impact on women themselves. Research (Paige, 1973) indicates that socio-cultural interpretations of biological processes, such as menstruation, are significant in the determination of responses to these processes. Thus, women as well as men learn that it is possible and maybe even expected that women experience some monthly distress. Again, the point is simply that we need to stop and consider if such interpretations affect performance criteria and evaluations.

What I have tried to do in these remarks is to raise some issues that need to be taken into account as Service Academies design follow-up studies and evaluations of male and female officers. These are particularly crucial as personnel become more diverse and integrated into a variety of non-traditional roles. In general, we need to take into account many of the attributions, stereotypes and assumptions which are operating in military as well as civilian settings

References

Acker, J., "Issues in Sociological Study of Women's Work," in Working Women, Stromberg and Harkness, eds., Palo Alto, Calif: Mayfield Publishers, 1978.

Binkin, M. and Bach, S., Women and the Military, Washington, D.C.: The Brookings Institute, 1977.

Calder, B., "An Attribution Theory of Leadership," in New Directions in Organizational Behavior, Starr and Salancik, eds., Chicago: St. Clair Press 1977.

Deaux, K., The Behavior of Men and Women, Monterey, Calif: Brooks/Cole, 1976.

Janowitz, M., The Professional Soldier, New York: McMillan Publishers, 1971.

Paige, K., "Women Learn to Sing the Menstrual Blues," in Psychology Today, September 1973, 41-46.

Phillips, D. and DeFleur, L., "Gender-Ascription and the Stereotyping of Deviants, Social Forces," (Forthcoming), 1980.

Rice, R.; Bender, L.; and Vitters, A., "Leader Sex, Follower Attitudes toward Women, and Leadership Effectiveness: A Laboratory Experiment," in Organizational Behavior and Human Performance, 25, 1980, 46-78.

Rustel, M., "The Khaki Collared Women of Company C," paper presented at American Sociological Association Meeting, Boston, 1979.

Stogdill, R., Handbook of Leadership: A Survey of Theory and Research, New York: Free Press, 1974.

BIGNESS IS IN THE EYE OF THE BEHOLDER

Stanley N. Roscoe
New Mexico State University

This paper is an integrated summary of a program sponsored partially by Ames Research Center, NASA, but mainly by the Air Force Office of Scientific Research. It was presented jointly by the author and a panel of former University of Illinois graduate students including Joyce H. Iavecchia, M.A.; Jan C. Hull, M.A.; Russell A. Benel, Ph.D.; Nicholas M. Simonelli, Ph.D.; and Valerie J. Gawron, Ph.D. The experiments encompass a variety of psychophysiological issues involving visual accommodation and judgments of size and distance as required by pilots in flight. Issues investigated included the size-distance invariance hypothesis, the moon illusion, night and empty-field myopia, the dark focus and its so-called Mandelbaum effect, the nature and locus of the accommodative stimulus, and possible relationships between accommodative responses and autonomic balance, personality, and age.

PROLOGUE

Thou flatterer! Do I not know beauty is altogether in the eye of the beholder, and that all persons do not see alike?

General Lew Wallace of Indiana, The Prince of India, 1893, p. 178

. . .

It (the moon illusion) is not due to physical causes outside the visual mechanism. It is not due to the greater brightness of the moon in elevation, when atmospheric haze is diminished. It depends on raising or lowering the eyes.

Edwin G. Boring of Harvard, American Journal of Physics, 1943, p. 874

. . .

When the eye is fixated on a point in space, that point is sharply focused on the retina. All points nearer or farther than the fixated point are blurred. The blur circles formed for points not in focus have diameters that vary ... as a function of the distance between the fixated point and the point not fixated.

The discrimination of the clearly focused image from the blurred image probably serves as a distance cue, although the process of discriminating must be quite complex and certainly has never been adequately investigated. Presumably such factors as the discrimination of detail and retinal image size combine with the subjects' reactions to provide the cue.

Clarence H. Graham of Columbia, Handbook of Experimental Psychology, Stevens (Ed.), 1951, pp. 882-883

BACKGROUND

A research program at the University of Illinois, sponsored by the Air Force Office of Scientific Research, started as an investigation of certain evidently related misperceptions experienced by pilots flying airplanes and airplane simulators by reference to periscopes and other dynamic image projection systems (Roscoe, Hasler, and Dougherty 1966; Palmer and Cronn 1973). Imaged objects such as airport runways appear smaller and farther away than objects subtending the same visual angles viewed directly; pilots making landing approaches tend to come in too high and land long and hard (Roscoe 1979). Our investigation has uncovered clues that have led us far afield from airplanes and airports and into the realm of basic psychophysical and psychophysiological research.

This two-year program descended from an abortive and unpublished effort at the University of Illinois in the early 1970s by a graduate student with the unlikely name of Gleason Snashall. When Gleason discovered the size-distance invariance hypothesis in the literature (Weintraub and Gardner 1970), he developed a computer program to show that an object subtending a given visual angle must appear smaller than lifesize if it appears nearer than it really is; hence, an airport scene imaged on a nearby screen (or viewed through a collimating field lens) must appear smaller than lifesize, thereby causing pilots to overshoot.

For Gleason the problem was solved, but his logic failed to explain the phenomenon to my satisfaction; indeed, Gleason made me mad, and I resolved to pursue the investigation personally at the earliest opportunity. With support from the National Aeronautics and Space Administration, Robert J. Randle and I conducted a series of studies at Ames Research Center during 1975 and 1976 (Roscoe, Olzak, and Randle 1976; Roscoe 1977; Roscoe and Benel 1978; Roscoe 1979; Randle, Roscoe, and Petitt in press). These studies all involved the automatic covert measurement of visual accommodation using an infrared tracking optometer developed for NASA by Cornsweet and Crane (1970).

The experiments ranged from measurements of the apparent sizes of discs subtending equal visual angles at different distances to the tendencies of pilots to undershoot or overshoot landing approaches when viewing dynamic computer-generated night visual scenes. There is a strong correlation between apparent size and visual accommodation distance and a weaker interaction between overshoot/undershoot judgments and accommodation to real images (but not to virtual images); other things being equal, the more distant the eyes accommodate, the larger an object of fixed angular size appears. However, the eyes do not obediently accommodate to the distances of foveally presented stimuli (as investigators often assume); in fact, eyes focus stimuli only well enough for the required discrimination.

APPARATUS

Concurrently with the Ames studies, equipment was being developed by Oskar Richter at the University of Illinois (under an earlier contract from the Air Force Office of Scientific Research) for use in the present program. This equipment included a viewing system colloquially referred to as "the moon machine." This device projects a collimated disc subtending a 0.67-degree visual angle onto a combining glass through which a subject can view any 45 x 45-degree stimulus scene (Figures 1 and 2). A comparison disc of variable diameter can be viewed alternatively by the insertion of a mirror on a sliding assembly, thereby allowing a subjective matching with the apparent size of the collimated "moon" seen against various backgrounds (à la Kaufman and Rock 1962).

Also alternatively, a subject can view the flowing speckle pattern produced by a laser optometer (of the type developed by Leibowitz and Hennessy 1975) superposed on any stimulus scene by means of a small combining glass immediately in front of the subject's eye. The operation of this system is described in greater detail by Iavecchia, Iavecchia, and Roscoe (1978) and by Benel (1979). It provides an absolute though subjective measure of visual accommodation in diopters (D), a scale inversely related to the distance to which the eye is focused (D = 1/distance in meters). The measurements obtained correlate highly with those of a polarized vernier optometer developed and used by Simonelli (1979a) later in the program.

Simonelli's optometer employs an old principle attributed to Scheiner (see Duke-Elder 1970, p. 155), described by Moses (1971), and more recently advanced by consultant Robert T. Hennessy (personal communication). In Simonelli's words, referring to Figure 3:

> Using two pairs of perpendicularly oriented polarizing filters, the retinal image of a viewed object—in this case, a horizontal bar—will split when the retina is not conjugate with the plane of that bar. Likewise, the image will be whole when the retina is conjugate with the bar. This is an application of the Scheiner principle, . . . whereby one image (here, one half of the bar) is directed through the upper half of the pupil, and another image (the other bar half) is directed through the lower half.
>
> This direction of bar halves through different portions of the pupil is accomplished by creating bar-segment images whose light rays are of different polarities (indicated in the figure by the direction of the parallel lines in the filters). The left half of the target bar, for instance, is vertically polarized. Such rays will pass through the upper portion of the next pair of filters (with some absorption loss), as the polarities of the light and filter are identical. These vertical rays, however, cannot pass through the horizontal filter below.

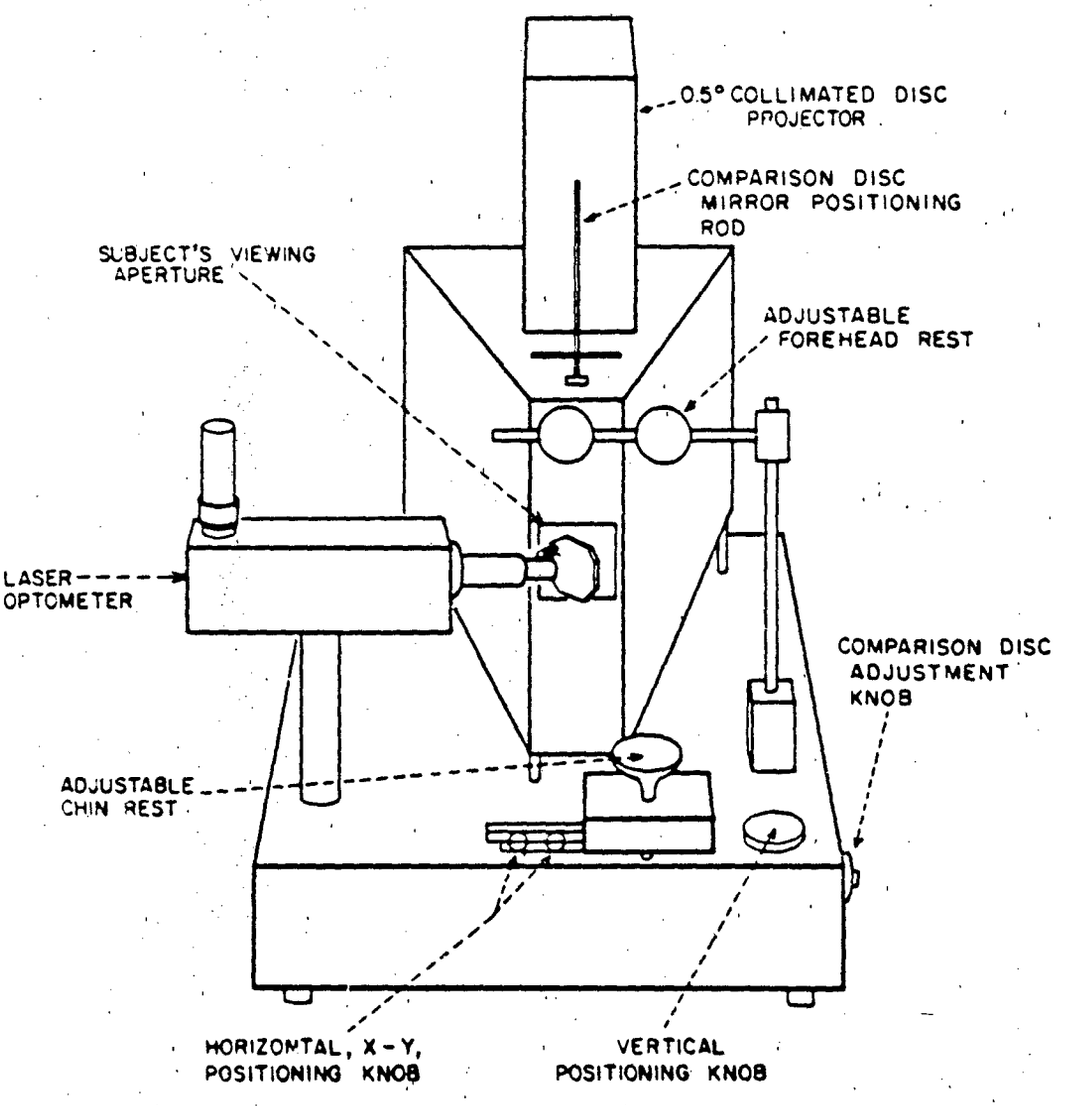


Figure 1. The Illinois "moon machine" and laser optometer used in experiments involving measurement and correlation of the apparent size of a simulated lunar disc and visual accommodation distance.

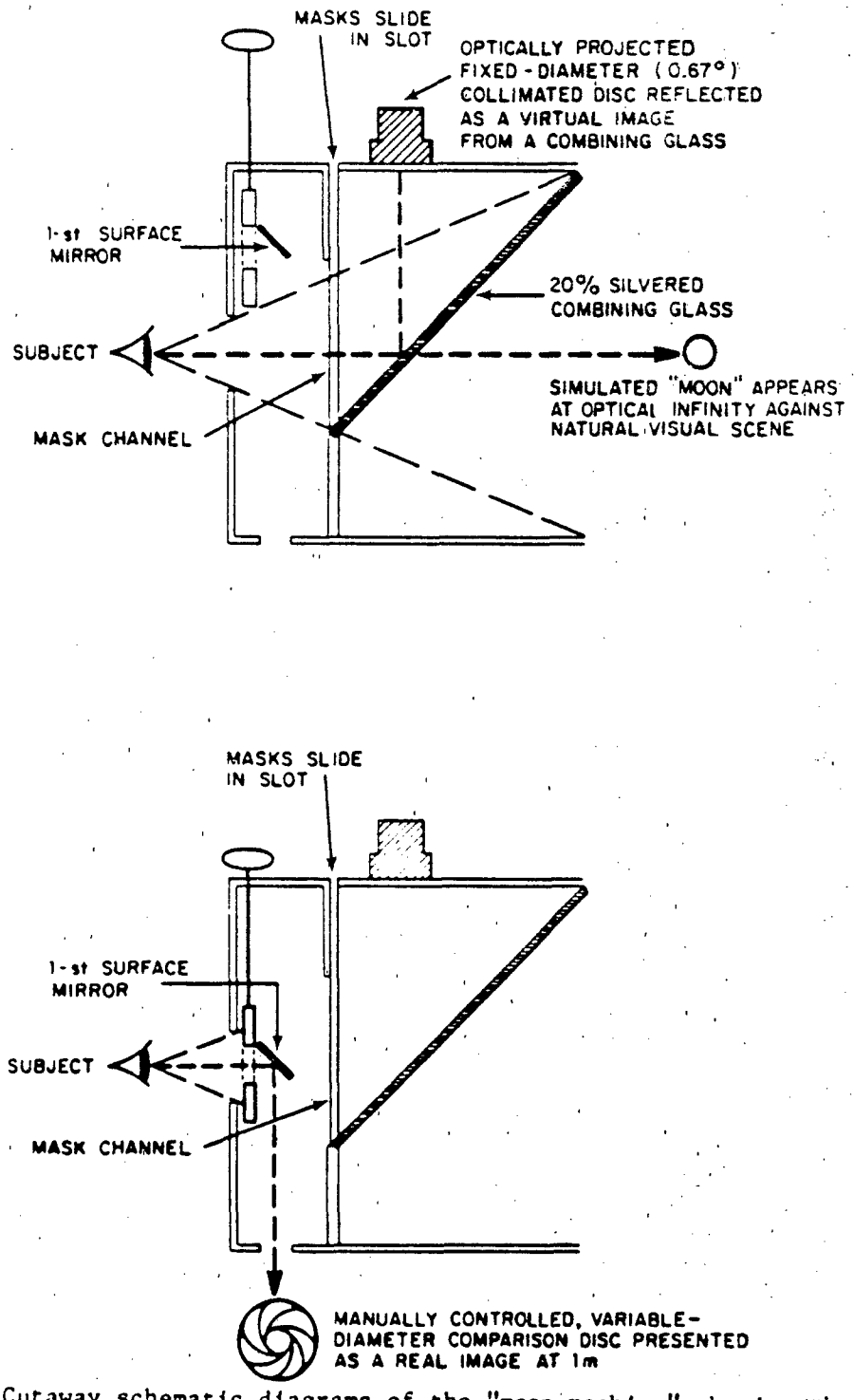


Figure 2. Cutaway schematic diagrams of the "moon machine" showing the presentation of the collimated lunar disc (above) and the variable-diameter uncollimated comparison disc (below).

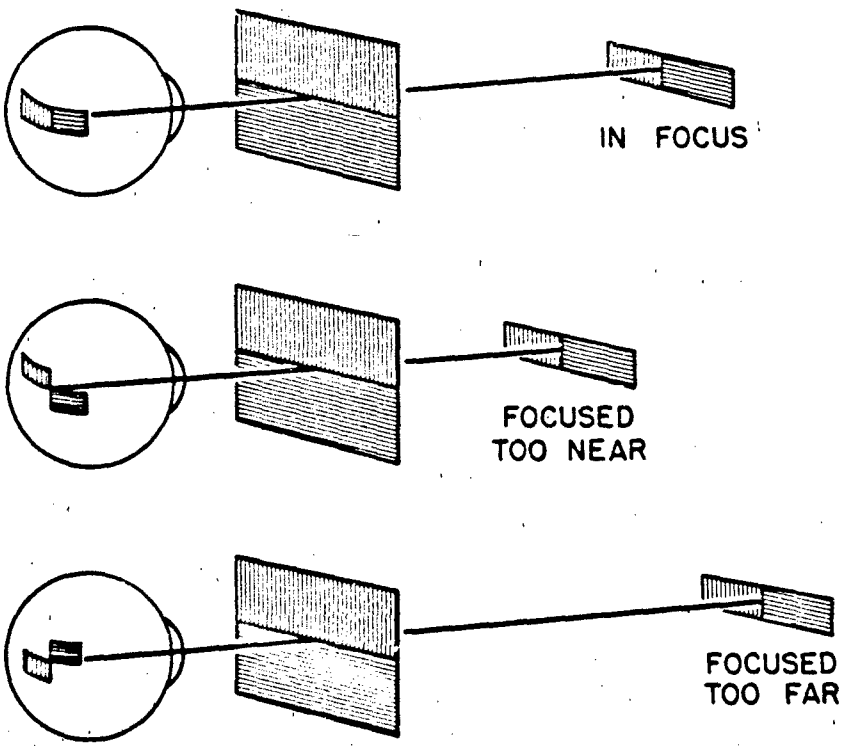


Figure 3. Diagram illustrating the use of orthogonally polarized filters to implement the Scheiner principle in Simonelli's polarized vernier optometer. An image polarized in one direction enters the eye through the upper half of the pupil; an image polarized orthogonally enters through the lower half.

Consequently, when this second pair of filters is aligned to "split" the pupil in half, the vertically polarized rays from the left portion of the target bar enter only the upper half of the pupil. Similarly, the image of the right half of the bar enters only the lower half of the pupil. When the eye is focused on the bar, both halves will "meet" at the retina and reform the whole bar. Moreover, one half will shift relative to the other when the eye is focused in front of or behind the stimulus bar. The amount and direction of the shift are related to the amount and direction of the focal error.

EXPERIMENTAL INVESTIGATION

Between 1977 and 1979, eight graduate students at the University of Illinois conducted seven experimental investigations, some involving more than one experiment. These studies are described in the following technical reports, the first three issued by the Department of Psychology, University of Illinois at Urbana-Champaign, and the last four by the Behavioral Engineering Laboratory, New Mexico State University:

Iavecchia, J. H., Iavecchia, H. P., & Roscoe, S. N. The moon illusion: Apparent size and visual accommodation distance (Tech. Rep. Eng Psy-78-4/AFOSR-78-3). Urbana-Champaign, IL: University of Illinois, Department of Psychology, 1978.

Benel, R. A., & Benel, D. C. R. Accommodation in untextured stimulus fields (Tech. Rep. Eng Psy-79-1/AFOSR-79-1). Urbana-Champaign, IL: University of Illinois, Department of Psychology, 1979.

Simonelli, N. M., & Roscoe, S. N. Apparent size and visual accommodation under day and night conditions (Tech. Rep. Eng Psy-79-3/AFOSR-79-3). Urbana-Champaign, IL: University of Illinois, Department of Psychology, 1979.

Benel, R. A. Visual accommodation, the Mandelbaum effect, and apparent size (Tech. Rep. BEL-79-1/AFOSR-79-5). Las Cruces, NM: New Mexico State University, Behavioral Engineering Laboratory, 1979.

Gawron, V. J. Eye accommodation, personality, and autonomic balance (Tech. Rep. BEL-79-2/AFOSR-79-6). Las Cruces, NM: New Mexico State University, 1979.

Simonelli, N. M. The dark focus of visual accommodation: Its existence, its measurement, its effects (Tech. Rep. BEL-79-3/AFOSR-79-7). Las Cruces, NM: New Mexico State University, Behavioral Engineering Laboratory, 1979.

Hull, J. C., Gill, R. T., & Roscoe, S. N. Locus of the stimulus to visual accommodation: Where in the world, or where in the eye? (Tech. Rep. BEL-79-5/AFOSR-79-9). Las Cruces, NM: New Mexico State University, Behavioral Engineering Laboratory, 1979.

An eighth technical report, issued by the University of Illinois, was based on data collected in 1976 by Lynn A. Olzak and Donna Miller at Ames Research Center:

Roscoe, S. N., & Benel, R. A. Is the eye smart or the brain forgiving? (Tech. Rep. Eng Psy-78-1/AFOSR-78-1). Urbana-Champaign, IL: University of Illinois, Department of Psychology, 1978.

A ninth report, issued by New Mexico State, described the principle, mechanization, and use of the polarized vernier optometer:

Simonelli, N. M. Polarized vernier optometer (Tech. Rep. BEL-79-4/AFOSR-79-8). Las Cruces, NM: New Mexico State University, Behavioral Engineering Laboratory, 1979.

The experimental investigations covered a broad spectrum of psychophysiological issues involving the measurement of visual accommodation and its correlation with various other dependent variables. The latter included judgments of apparent size, visual acuity discriminations, performance on a short-term memory task, physiological measures of autonomic balance, scores on a personality test of introversion-extraversion, and responses to a personal inventory questionnaire. Psychophysiological issues investigated included the size-distance invariance hypothesis, the moon illusion, night and empty-field myopia, the dark focus and its so-called Mandelbaum effect, the nature and locus of the accommodative stimulus, and possible relationships among accommodative responses, autonomic balance, and personality.

Apparent Size and Accommodation to Visible Texture

The Ames studies had established a reliable correlation between apparent size and accommodation to targets at distances up to 4 meters. Pilots flying airplanes by contact visual reference view objects at far greater distances. A convenient way to study perceptual and accommodative responses to distant scenes is to use a technique developed by Kaufman and Rock (1962) to quantify the moon illusion. By superposing a collimated disc of light on a natural outdoor scene and providing an adjustable-diameter disc nearby, a surprisingly accurate estimate of the apparent size of a distant object ("the moon") can be obtained. This technique was used by Joyce and Helene Iavecchia.

In their first experiment, the Iavecchia sisters had subjects view the collimated moon against the scenes from corresponding windows of the third through the eighth floors of the psychology building looking eastward across the Urbana-Champaign campus. As shown in Figure 4, the apparent size of the moon increased from the third through the sixth floors and decreased thereafter. Although accommodation was not measured in this experiment, the foveal textural stimuli visible from the various elevations appeared at distances ranging from about 30 meters (the roof of a nearby sorority house) to more than 1000 meters (trees and buildings across the Urbana campus).

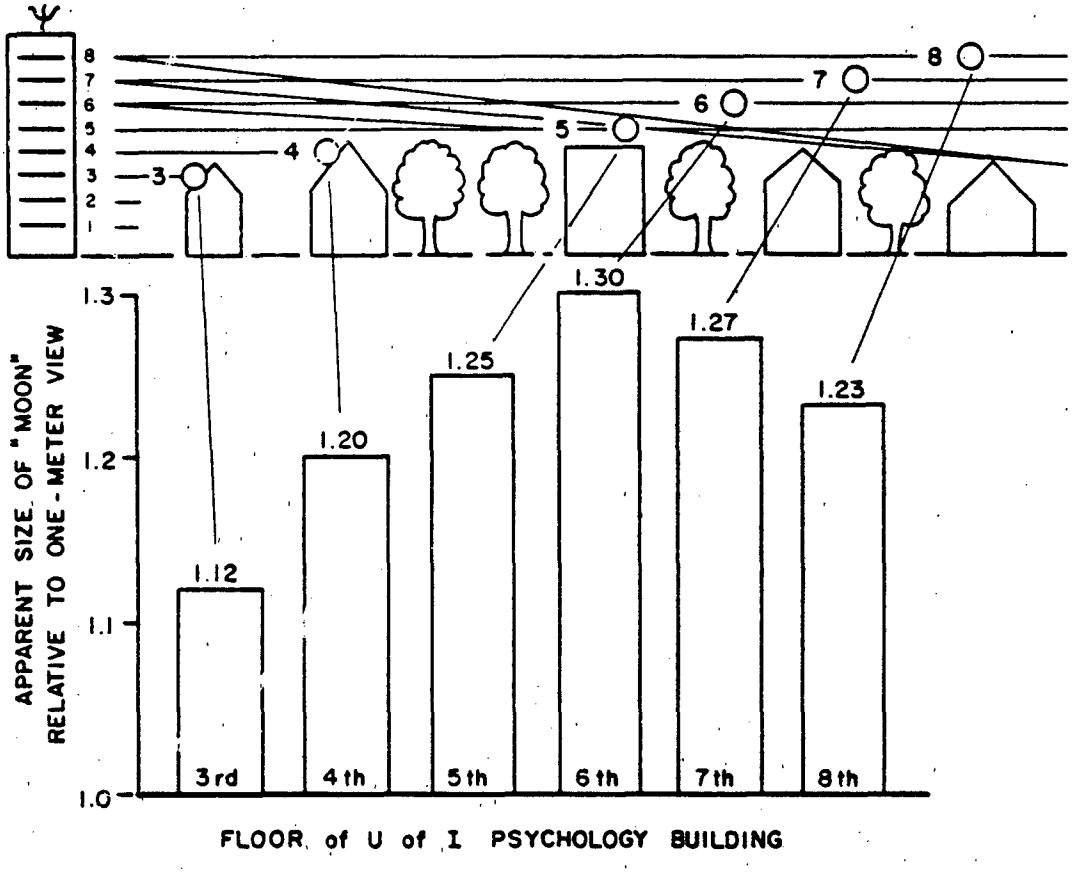


Figure 4. Apparent diameter of the simulated moon, when viewed from different floors of the psychology building, expressed as ratios of its apparent diameter when projected onto a newspaper at one meter.

In a second experiment from the sixth-floor elevation, the distance and angular depression of visible texture were systematically manipulated by use of a series of masks, as illustrated in Figure 5. In this experiment visual accommodation was measured with a laser optometer and correlated with subjective judgments of the apparent size of the collimated moon projected just above the distant horizon. A systematic relationship ($r \approx .9$) was found between accommodation distance and the apparent size of the moon, a finding quite consistent with the hypothesis that accommodation varies with the locus of textural stimuli all of which are well beyond what is nominally though of as "optical infinity," as shown in Figure 6.

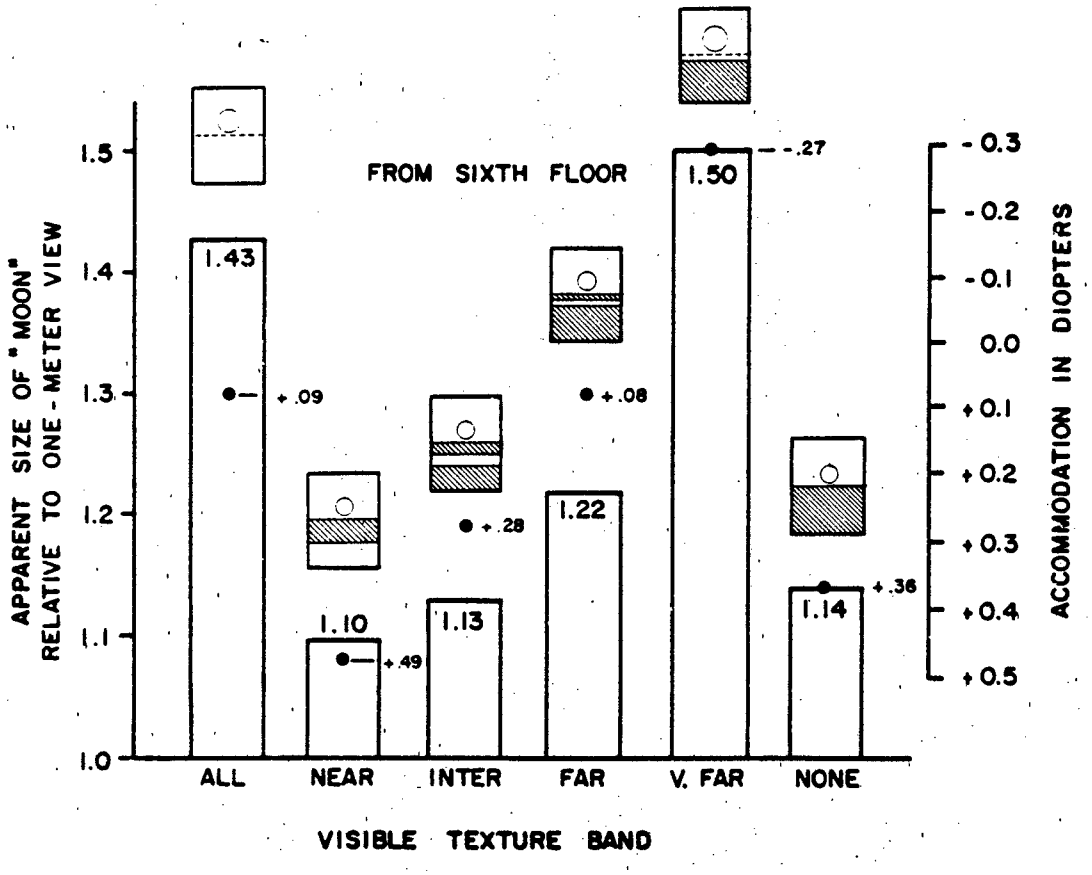


Figure 5. Apparent diameter of the moon, expressed as ratios of its apparent diameter when projected onto a newspaper at one meter, and associated visual accommodation distances, expressed in diopters, when various horizontal bands of texture are visible in the natural scene from the sixth floor (adapted from Iavecchia, et al., 1978).

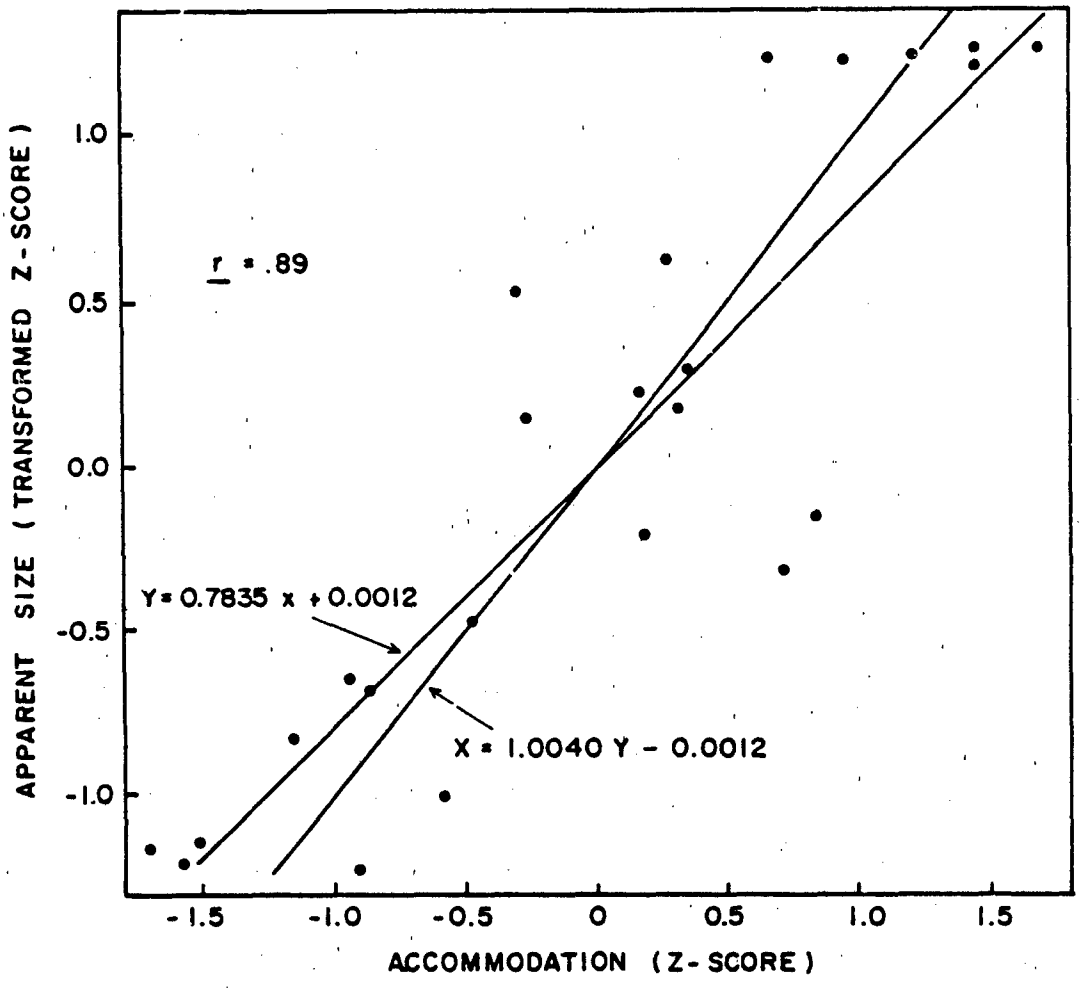


Figure 6. Scatter plot of the relationship between the apparent size of the moon, expressed as a transformed Z-score (specifically, 1.23 Diameter - 0.47 Diameter^2 + 0.47) and visual accommodation distance, expressed as a Z-score (Iavecchia, et al., 1978).

The Iavecchia experiments established the relationship between apparent size and accommodation for distances far beyond those investigated in the earlier Ames studies, but they also raised additional questions that called for further investigation. The nature and locus of the accommodative stimulus were not untangled and in fact were still totally confounded. Nicholas Simonelli investigated the effects of the reduced nature of textural stimuli under night as opposed to daylight viewing conditions (Simonelli and Roscoe 1979), and Jan Hull and Richard Gill (with assistance from Thomas Bolitho) investigated the locus of the accommodative stimulus (Hull, Gill, and Roscoe 1979).

There is conclusive evidence of functional myopia in the absence of a proximal textural gradient as in flight (empty-field myopia) or the reduced gradient in the relative darkness of night. Simonelli conducted an experiment similar to the Iavecchia manipulations but from the roof of the eight-story psychology building both in the daylight and at night. In the daylight the nearest textural stimulus appeared at about 100 meters with "empty space" intervening. At night the lights of the city were visible at an even greater distance. The masking manipulations had relatively little effect on the myopic accommodative responses and apparent size judgments, but the correlations between accommodation and apparent size were virtually identical to those obtained by the Iavecchias in daylight ($r \approx .9$) and slightly lower at night ($r \approx .7$).

Where in the World or Where in the Eye?

In the Iavecchia experiments the distance to dominant textural stimuli and their retinal locus were confounded. In the first experiment (in which accommodation was not measured), the distance to visible textural stimuli increased as the views progressed from the third through sixth floors, and the apparent size of the moon increased accordingly. From the sixth through eighth floors the dominant visible texture appeared farther and farther below the foveally presented collimated moon, which decreased in apparent size accordingly. In the second experiment (in which accommodation was measured), masks were used to obscure or reveal horizontal bands of texture whose absolute distance and angular depression from the foveal stimulus covaried inversely.

The subsequent experiments by Hull and Gill involved independent manipulations of retinal locus and absolute distance of stimuli, thereby helping to untangle the previous confounding but with a surprising result requiring further untangling. In their first experiment (employing three masks exposing near/low, intermediate, and far/high textural bands, respectively), the sixth-floor view across the Urbana campus was photographed (the same view used in the second Iavecchia experiment), and the color slide was rear-projected on a screen viewed by subjects through a large (25-inch diameter) collimating field lens immediately in front of the moon machine, as shown in Figure 7.

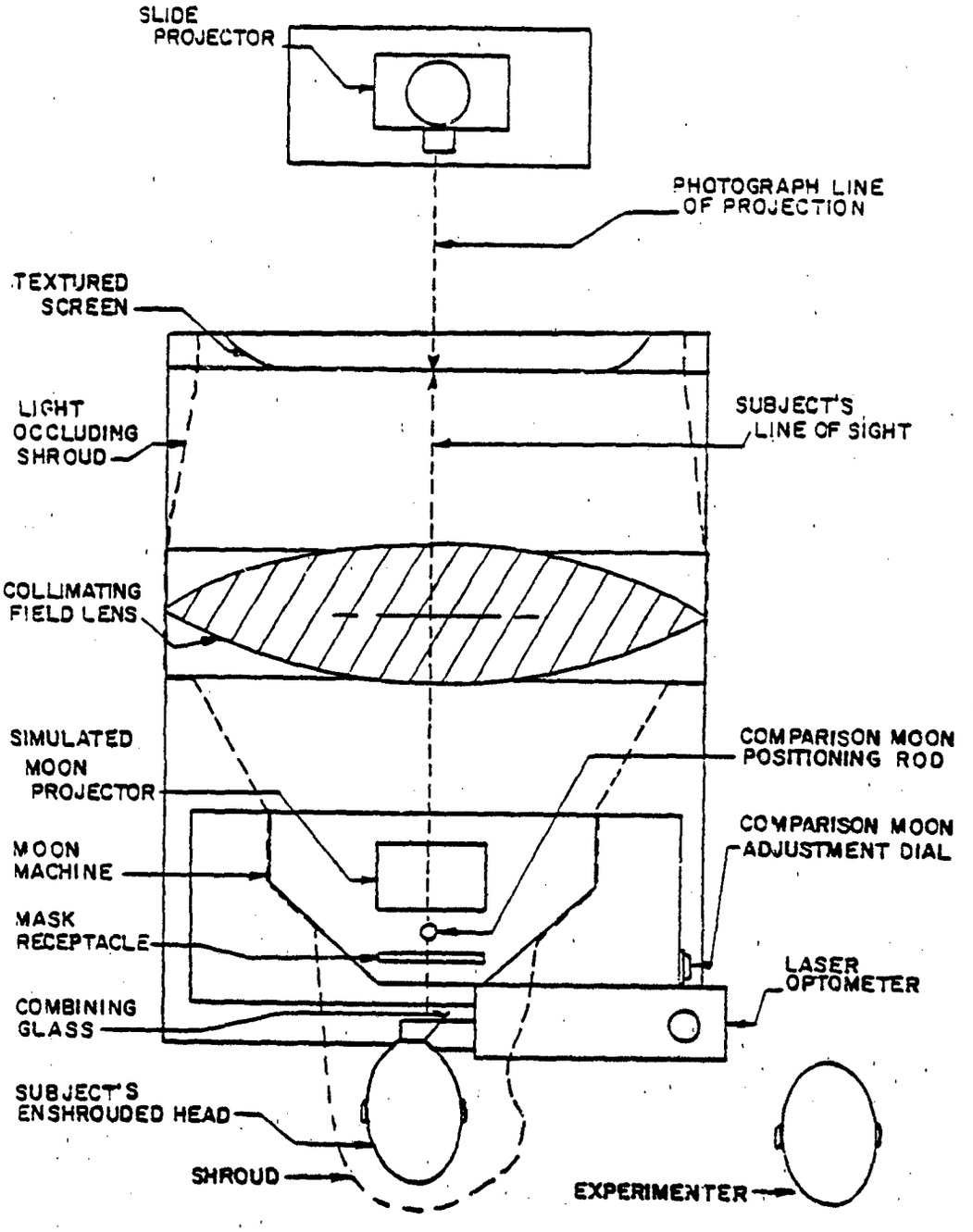


Figure 7. Arrangement of apparatus used by Hull and Gill. Subjects made judgments of the apparent size of the simulated moon when combined optically with collimated projections of a color photograph of the sixth-floor view of the Urbana campus through various masks. Associated visual accommodation distances were measured with a laser optometer.

This arrangement provided an image of the entire scene from "optical infinity," thereby allowing manipulation of the locus of retinal stimulation while holding "optical distance" constant. To assess the stimulus value of a two-dimensional photographic image of a three-dimensional real-world scene, a second set of accommodation and apparent-size measurements were made while the subjects viewed the simulated moon against the backlighted screen providing collimated texture through the various masks but no scenic image. To allow direct comparison of responses in these conditions to those in the second Iavecchia experiment, measurements of responses to the real-world sixth-floor scene were repeated for the eight subjects in this study.

Once again the mask manipulations resulted in differential shifts in accommodation, this time measured relative to the individuals' dark focus distances, and corresponding shifts in the apparent size of the moon. However, as shown in Figure 8, the absolute accommodation levels to either of the collimated two-dimensional stimulus views were highly myopic relative to those from the three-dimensional out-the-window scenes, and the magnitude of the moon illusion varied accordingly. Figure 8 shows the mean apparent size of the moon plotted against the mean accommodation level for the eight subjects in response to each of the three textural bands (masks) for each of the three scenes (views).

The correlation between these two sets of means is 0.97. Despite the orderliness of this relationship, inspection of the 72 individual points (3 masks by 3 views for each of 8 subjects) showed that apparent size and accommodation were not related in a purely linear fashion. As in the Iavecchia study, the apparent diameter of the moon increased disproportionately with outward accommodation expressed in diopters. Evidently the dioptric scale does not represent equal psychophysical units. As was done by the Iavecchias, the apparent-size scale was transformed, but in this case the reciprocal of the square of the linear dimension (the diameter of the moon) yielded the best-fitting linear relationship. This suggests that apparent area is a linear function of the focal distance of accommodation.

So far, so good. Visual accommodation and apparent size (when suitably transformed) bear a strong linear relationship when the retinal locus of a textured accommodative stimulus is varied systematically relative to a foveally presented untextured target object (the moon). This relationship holds even though the responses to collimated two-dimensional scenes are much nearer (in diopters) and smaller (in degrees) than corresponding responses to three-dimensional real-world scenes. But what happens when the absolute distance to visible real-world texture is varied while holding the retinal locus of the textural stimulus approximately constant?

To investigate this question, Hull and Gill partially replicated the first Iavecchia experiment, this time measuring accommodation. Using the intermediate mask that revealed a horizontal band depressed from the line-of-sight, subjects viewed the simulated moon above rooftops and treetops at successively increasing distances from the fifth-, sixth-, seventh-, and eighth-floor views across the Urbana campus, as illustrated in Figure 9. The mean size judgments and accommodative responses to this manipulation

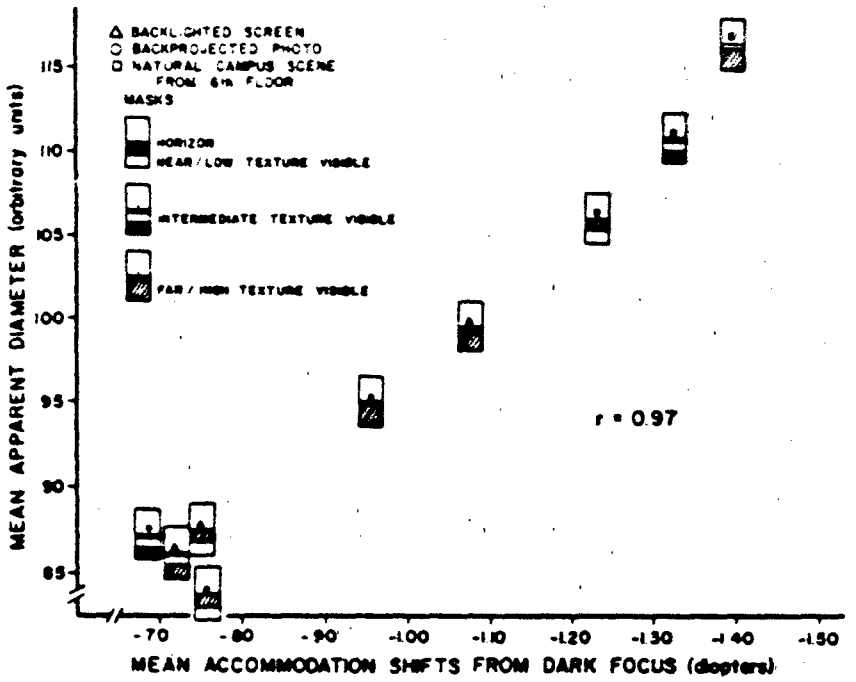


Figure 8. Mean apparent diameter of the simulated moon (in arbitrary units) as a function of the mean accommodation shift from individual dark focus levels for eight subjects viewing three textural bands (masks) for each of three scenes (views).

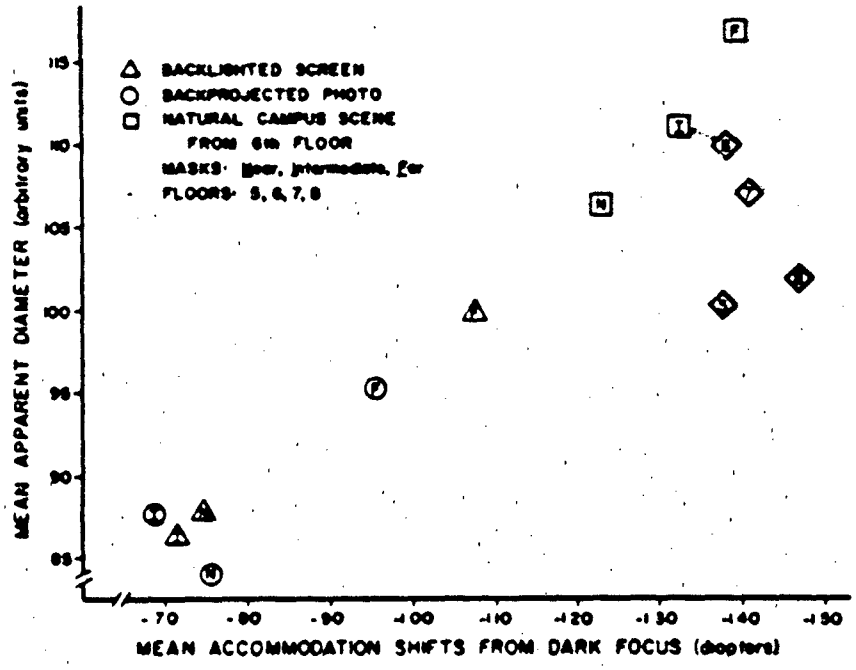


Figure 9. Mean apparent diameter of the simulated moon (in arbitrary units) plotted against mean accommodation shifts from individual dark focus levels for eight subjects viewing the campus vista through the intermediate mask from various floors of the psychology building. (Data points from Figure 8 are repeated for comparison.)

are superposed on the corresponding values from Figure 8 for the previous experiment by Hull and Gill. Neither apparent size nor accommodation varied reliably although apparent size tended to increase from the fifth- to sixth-floor views and decrease thereafter as observed in the first Iavecchia experiment.

On the surface it would seem that the retinal locus of a textured stimulus is the primary determinant of the accommodation response of the eye, but what then determines apparent size and distance? Clearly a more systematic and precise manipulation of the retinal locus and absolute distance of more uniform textural gradients is required to untangle this puzzle. Evidently from the earlier experiments eye accommodation is involved in some way in judgments of apparent size, but the nature of the involvement is far from clear. Accommodation must interact in some complex manner with other variables as yet not fully identified. Study of the problem continued.

Accommodation in Untextured Stimulus Fields

The background studies at Ames Research Center and all of the Illinois studies discussed so far involved textured visual stimuli, in most cases real-world vistas or photographic projections thereof. To understand the contributions of texture in determining the accommodative reflex, Russell and Denise Benel (1979) performed a series of experiments involving untextured stimulus fields. In their words:

> When textural cues are reduced through lowered illumination and contrast, night and empty-field or space myopia occur. Instrument myopia is typically attributed to viewing through small apertures. These myopias are referred to as anomalous because, contrary to the classical view, accommodation for near images occurs in the absence of...[textured stimuli]... (Leibowitz and Owens, 1978). An alternative theory of accommodation proposes that there is dual control of the ciliary muscle and the intermediate state reflects a passive return to a neutral balance point between these opposing systems. Although this alternative theory has been frequently proposed and subsequently denied, anatomical and physiological evidence has mounted making it difficult to reject (see Cogan 1937 for a review of the early evidence or Benel 1979 for an updated review).

Although there were reliable mean differences in accommodation to the various untextured stimulus manipulations employed by the Benels, the measured responses in all cases departed little from the individual subjects' resting accommodation distance or dark focus. As has been reported by many investigators, in the absence of adequate textural stimuli at distances other than the dark focus, little accommodation occurs, either inward or outward. The correlation between accommodation to any untextured stimulus and the dark focus is extremely high, typically in excess of 0.9.

Visual Accommodation, the Mandelbaum Effect, and Apparent Size

Russell Benel pursued the implications of the evident relationships between the large differences among the dark focus distances of different individuals and their judgments of apparent size as influenced by textural patterns of varying spatial frequency and acutance at varying distances. His studies involved the phenomenon that has become known as the "Mandelbaum effect." Mandelbaum (1960) observed that he and several of his friends were unable to read a sign from inside a screened porch although it was clearly legible when viewed directly. He also noted that the sign could be read through the screen if one moved closer or farther from it or moved the head from side-to-side.

Mandelbaum concluded that he and his friends were accommodating involuntarily to the screen rather than the sign, and Owens (1979) has subsequently demonstrated that the critical distance from a screen at which the effect occurs is the individual's dark focus distance. Benel proceeded to review the vast literature that might have a bearing on this phenomenon including anatomical and neurophysiological considerations, the long controversy over the single versus dual innervation by the parasympathetic and sympathetic branches of the autonomic nervous system (see Cogan 1937), and the so-called anomalous myopias--night, empty-field, and instrument. In Benel's words:

> Recent evidence has made it clear that earlier conceptions of visual functioning bear reexamination. Roscoe and Benel (1978) have noted two misconceptions that have misdirected psychologists for more than a century. The first concerns the misbelief that the eye's relaxed accommodation distance is at the far point, for the emmetrope at "optical infinity." This legacy has been passed down from Helmholtz (1867/1962, vol. 1, p. 360) who declared "when it [the eye] is focused for the far point, ... accommodation, therefore, is relaxed." Concomitant with this view is a single innervation theory of control of the ciliary muscle. Frequently, belief in single innervation obscured the need for verification of the far resting point and vice versa.
>
> The second closely related misconception has been the belief that the eye reflexively accommodates accurately to the distance of an object present in foveal vision. This latter belief is often implicitly assumed to hold in laboratory experiments on visual sensation and perception. The importance of these topics is apparent to psychologists because of their historical concern for the role of oculomotor adjustments in space perception (Baird, 1970). These oculomotor adjustments represent the initial response to distance and determine the clarity of the retinal image. This in turn, has a fundamental influence on perception and on the information derived from the stimulus. (Benel, 1979, pp. 1-2)

Benel then drew upon Hoffman, as translated by Ogle (1950, p. 10):

> ...Look with one eye, while the other is closed, at a window several meters away. Then hold one finger so close in front of the active eye that you have to accommodate on it with difficulty. As soon as this is done, the window shrinks and seems smaller than when one observes it without the effort of accommodation. Of course, a measuring rod behaves in precisely the same way if it is applied to the window at that time. Thus, the objective size of the window gives us no information as to the subjective size, either of the measuring rod or of the object--the window--that it measures. The spatial extent of objects does not give us any standard for the size of subjective, visual objects. (Benel, 1979, pp. 2-3)

Picking up on the work of D. Alfred (Fred) Owens at Pennsylvania State University (now at Franklin and Marshall College), Benel did four experiments that comprised his doctoral research. Owens had hypothesized that stimulus characteristics necessary to draw accommodation away from the dark focus include textural contrast peaks between the spatial frequencies of 5 and 14 cycles per degree of visual angle; any target containing only spatial frequencies outside the observer's optimum sensitivity range should not be an effective stimulus for accommodation. Either low-frequency targets, such as gross, blurred shapes, or high-frequency targets, such as small, sharp fixation points, would be poor accommodative stimuli.

Owens also proposed quantifying the effects of stimuli of varying spatial frequencies by optically varying their distances and simultaneously measuring accommodation. Benel arranged to do this and also to measure the apparent size of a foveally presented collimated disc as a function of such manipulations. But first it was necessary to validate the appropriateness of the slope of the regression of accommodation on stimulus distance as a psychophysical (as opposed to a physical) index of stimulus adequacy. Textural stimuli that ranged from sharply imaged gratings to grossly blurred images of the same targets were presented at varying optical distances.

An Index of Stimulus Adequacy

As illustrated in Figure 10, stimuli were positioned in the collimated portions of the optical channel between lenses L3 and L4, and L5 and L6, respectively. Movement of the stimuli within each channel varied the optical distances independently. Lenses L3 through L6 are of 180 mm focal length yielding a maximum dioptric power for each channel of 5.56 D and equal magnification within each channel. The diameter of the circular stimulus field subtended a visual angle of 12 degrees. The size was limited by a field stop (FS2) placed at -5.56 D (beyond optical infinity). The field stop provided a severely out-of-focus edge image that would not act as an accommodative stimulus.

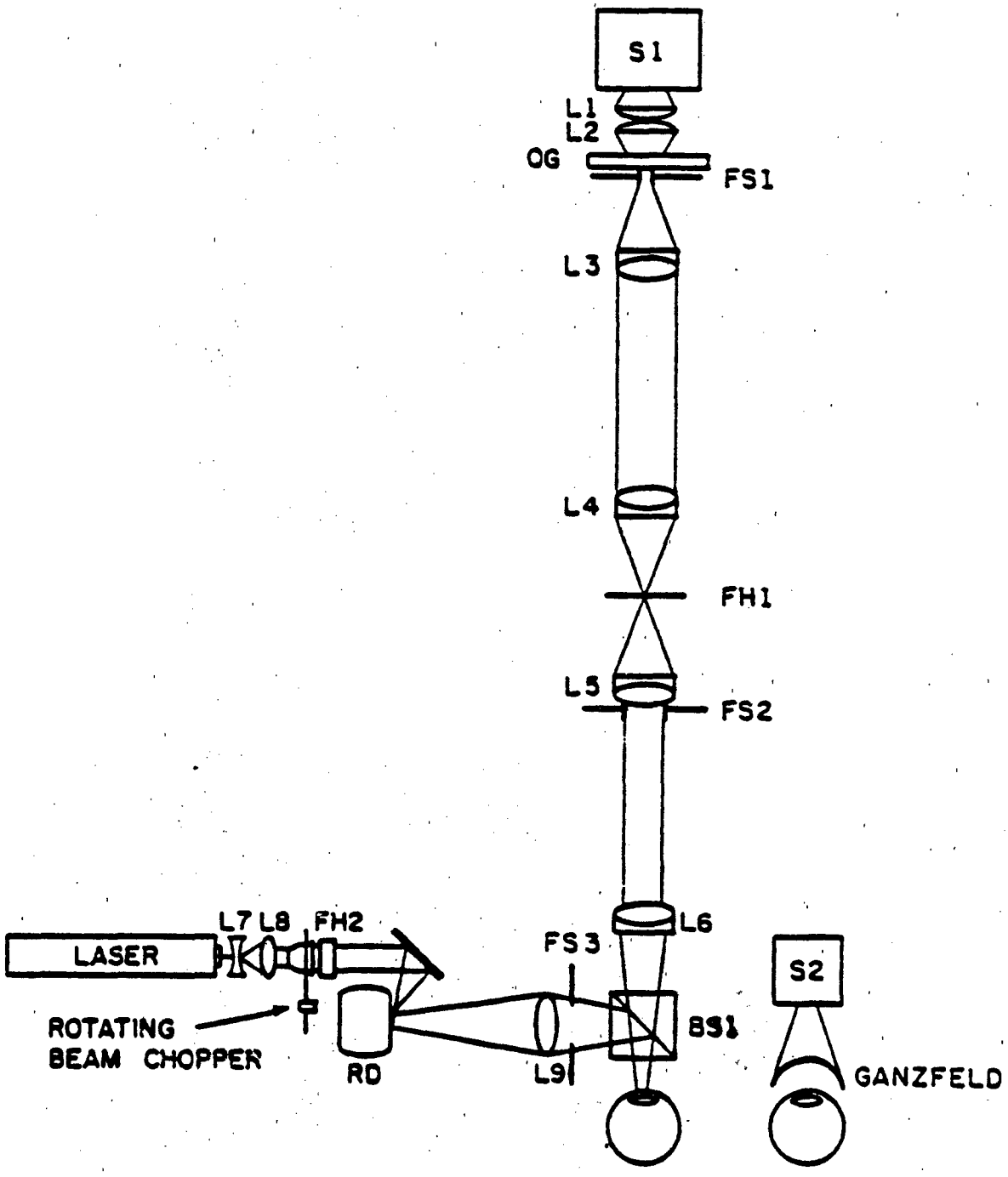


Figure 10. Schematic diagram of the Maxwellian viewing system and the accommodation measurement apparatus used by Benel (1979).

Results of Benel's first experiment are summarized in Figure 11. Accommodative responses to Screen 1 (the high-contrast grating) were reasonably accurate, approaching a slope of 1.0, throughout the range of stimulus presentation distances. With successively defocused gratings (Screens 2, 3, and 4), accommodation lapsed toward the dark focus, and the regression slopes flattened accordingly. Comparison of the accommodative regression slopes and the objective changes in stimulus characteristics (percent contrast) supports the validity of this metric as a functional (psychophysical) description of stimulus adequacy.

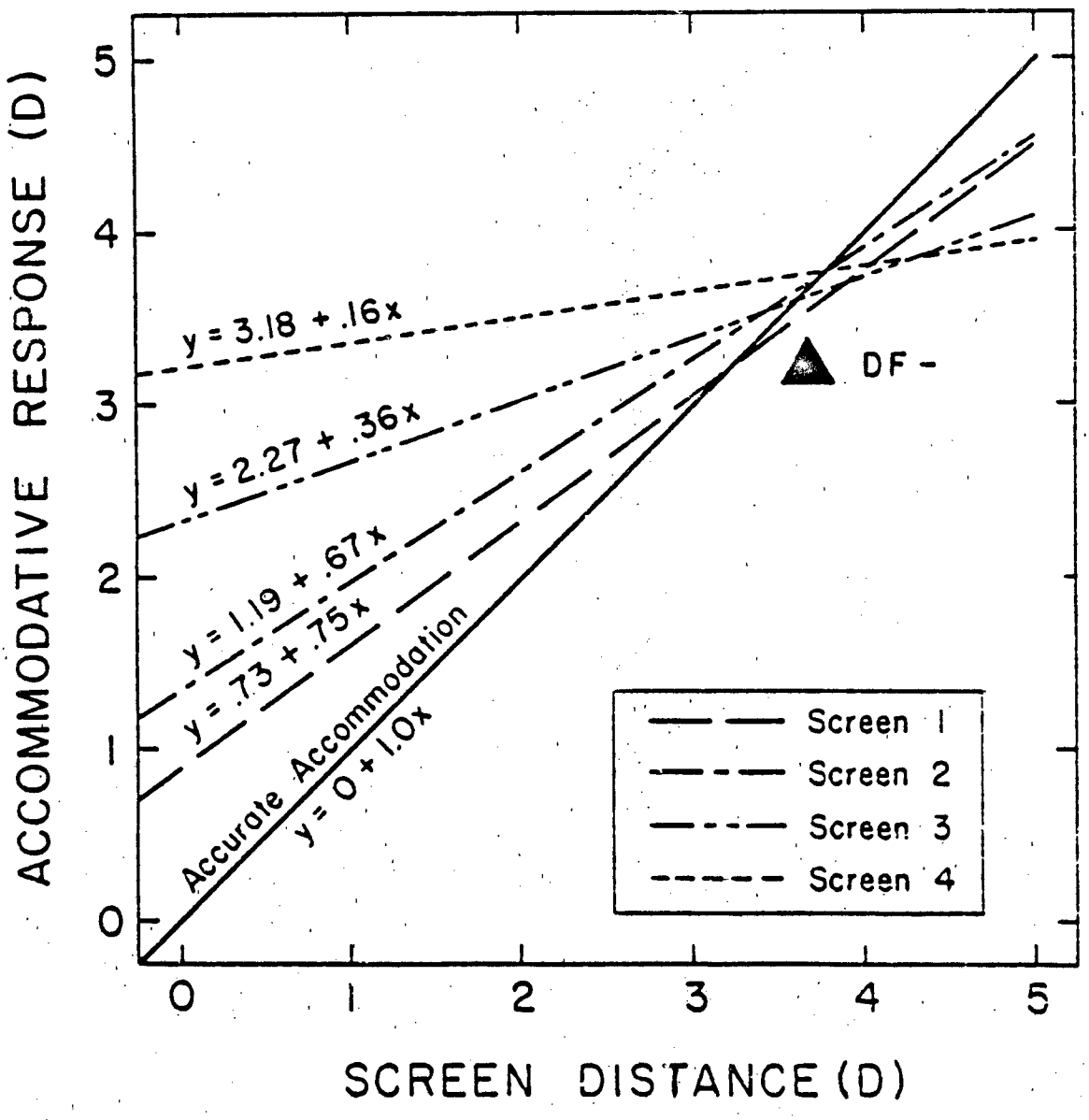


Figure 11. Regression of accommodation on stimulus presentation distance for each of four gratings varying in contrast ratio from high (Screen 1) to low (Screen 4) tested by Benel (1979).

Refining the Mandelbaum Effect

In his second experiment, Benel investigated the possibility, noted by Owens, that changes in the retinal image with changes in the observer's position could have confounded Mandelbaum's findings. The four screens of varying contrast were presented at varying optical distances while subjects attempted to resolve high-contrast acuity targets (matrices of Snellen Es of varying orientation) also presented at varying distances. The effects on accommodation of target distance, screen distance, and screen contrast and their first-order interactions were all reliable. Target distance accounted for 61 percent of the response variance, and the various effects of screens, screen distances, and their interactions accounted for an additional 20 percent.

For the two high-contrast screens (1 and 2) the effect of screen position on accommodation was quite evident. As screen position approached the individual's dark focus, the screen tended to draw accommodation away from the targets' distance. Under certain circumstances the high-contrast screens provided more potent stimuli than the matrix of Es. For example, when the target was at 3.75 D and either Screen 1 or 2 was more distant, accommodation was drawn outward. However, neither Screen 3 or 4 exerted a similar influence.

Apparent Size and Accommodation

Hoffman's observation, quoted earlier, indicates that inward shifts in accommodation are accompanied by reductions in apparent size. As Hoffman pointed out, objective measurement of such shifts is difficult because any measuring standard (visible scale) presented with the object to be judged changes proportionately. Under such circumstances, the ostensibly objective size will remain constant despite the subjective impression of a shrunken object and scale. The converse was not discussed explicitly by Hoffman but follows logically. An outward shift in accommodation should be accompanied by an increase in size of an object and its surroundings, as demonstrated successively by the Iavecchias, Simonelli, and Hull and Gill.

In these earlier experiments at the University of Illinois, stimuli presented included irregular and unquantified textural gradients, usually in natural vistas, that allowed various types of experimental confounding in the manipulation of accommodation. If accommodation were induced by targets and competing screens of quantitatively known stimulus adequacy, the relationships between induced accommodation distances and apparent size might be purified. Calling merely for subjective judgments of "smaller," "larger," or "no change," Benel inserted screens at various optical distances as subjects fixated targets, also set at various distances. The coincidence of "smaller" judgments with inward accommodation shifts and "larger" judgments with outward shifts was statistically reliable.

Apparent Size and the Mandelbaum Effect

In all experiments involving the Mandelbaum effect as such, investigators other than Benel have been concerned mainly with measurable decrements in visual acuity associated with the inappropriate accommodation induced. To quantify shifts in apparent size with the Mandelbaum effect of varying screen distances, Benel set up a fourth experiment. Using the moon machine in an arrangement similar to that used by the Iavecchias looking east from the fifth floor of the psychology building, Benel inserted a black fiberglass windowscreen at varying distances between the subject and the recently cleaned window. Nothing was changed in the subject's view of the campus vista except the position of the screen.

As illustrated in Figure 12, reliable shifts in apparent size of the collimated moon just above the distant horizon (as indicated by the subjects' settings of the uncollimated comparison moon at one meter) were associated with the accommodation shifts induced by the screen. The correlation of -0.96 between mean judgments of size and mean accommodation levels for the 12 subjects in the six conditions was slightly higher than observed in the previous experiments involving additional stimulus variations. Approximately 50 percent of the total variance in size judgments (including subject variance) was accounted for the the inappropriate shifts in accommodation.

This finding strongly suggests a causal relationship, but whether causal or not, the practical lesson is clear: beware the dirty windshield and the head-up flight display and be cautious at night.

The Dark Focus of Accommodation: Its Existence, Its Measurement, Its Effects

In view of the emerging relationships among the dark focus, myopia and hyperopia, and perceptual responses, it was time to investigate the effects of subject populations on experimental findings. Nicholas Simonelli proposed a dissertation project that his doctoral committee reluctantly approved as a "somewhat pedestrian undertaking," in the words of Professor Lloyd Humphreys, Dean of Liberal Arts and Sciences at Illinois. But Simonelli's dissertation was hardly pedestrian. From his scholarly, entertaining, and insightful analysis of relevant literature one might suspect that what we think we know about perception is based largely on the eyeballs of college sophomores, the most readily available and near-sighted human subjects in the world.

Compared to the professional pilots tested in the studies at Ames Research Center, undergraduates in universities with stiff academic requirements are, on average, myopic to a degree exceeded only by psychology graduate students and their spouses, another popular source of subjects. The mean dark focus distances of the student subjects in the earlier studies by Simonelli, the Benels, and Hull and Gill were all well within one meter, as has typically been found in the many studies by Leibowitz and his students at Penn State. This has been true despite the usual requirement that subjects exhibit normal visual acuity (often set at 20/25 rather than 20/20 and occasionally with optical corrections).

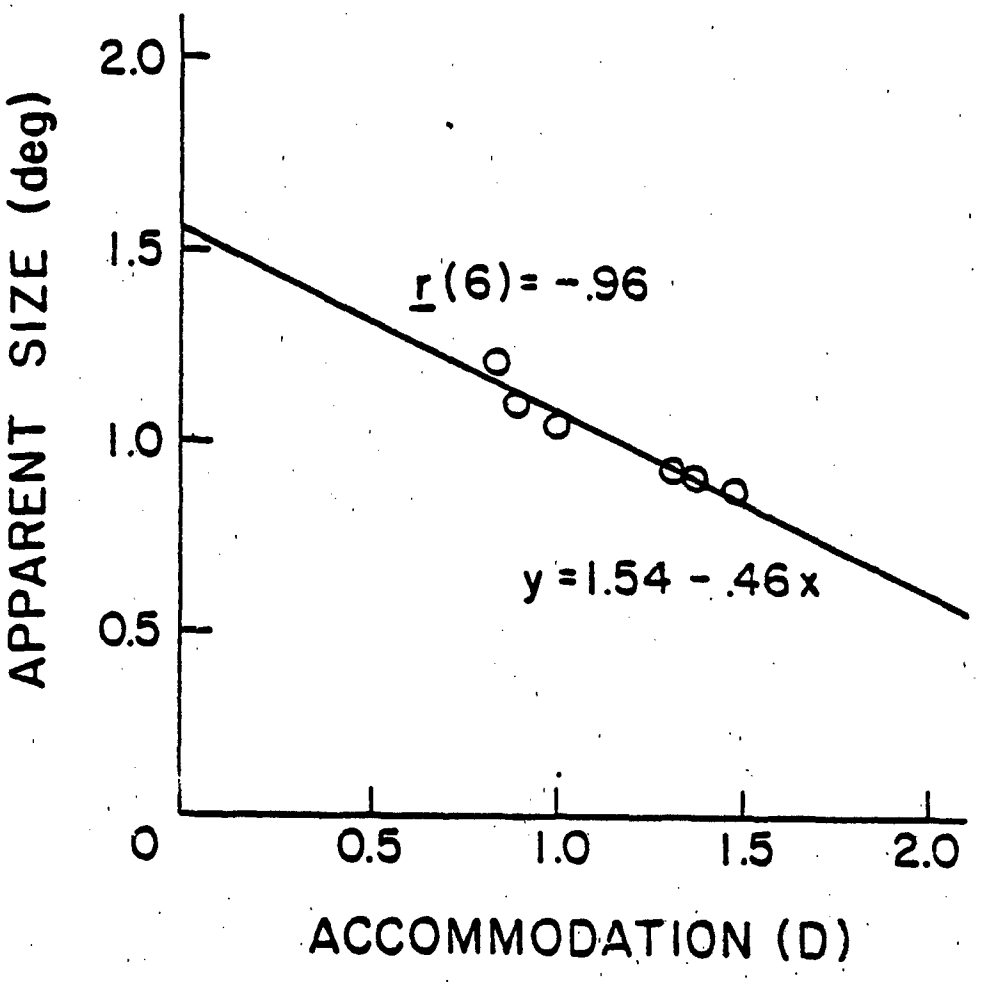


Figure 12. Apparent diameter of the collimated horizon moon over the campus vista as a function of inappropriate accommodation induced by a black mesh windowscreen interposed at optical distances of 3.0, 2.25, 1.5, and 0.75 D from the subject's eye position. Each point represents the mean size and accommodation values for 12 observers in the six viewing conditions including the four screen distances, the unscreened view, and the collimated moon with the outside view obscured. (Benel, 1979)

To study his main experimental questions, Simonelli first developed a new optometer, described earlier, that was simpler and easier to operate than the laser optometers used in the earlier experiments at Illinois. It also proved to yield slightly more repeatable measures of accommodation, but its chief advantages were its reliability of operation and its ease of explanation and use in the testing of large numbers of experimentally naive subjects, military trainees at Chanute Air Force Base as well as college students. The results of Simonelli's two main experiments using this device call for a critical reappraisal of the generality of many accepted findings in the vast literature on visual perception.

The Sampling of Eyeballs

In all, Simonelli tested 301 subjects ranging in age from 17 to 67. Of these, 114 were psychology students, and 154 were Air Force recruits of comparable ages (253 of the 268 were between 17 and 22); 33 subjects distributed over the range from 28 to 67 were neither students nor recruits. In addition to taking near point, dark focus, and far point measurements, Simonelli introduced the term "relative dark focus," the difference in diopters between an individual's dark focus and far point, the distribution of which is shown in Figure 13 for all 301 subjects combined. Separate statistics for the students and recruits are given in Table 1.

Although students and recruits did not differ reliably in terms of near point or relative dark focus on average, that was the end of the similarities. The students were almost 1.5 D more myopic than the recruits (far-point comparison), and their measured dark-focus distances were also 1.5 D closer (in centimeters, 37 versus 84). In Simonelli's words:

> These differences, especially the far points, will come as no surprise. Students are typically thought of as having poor vision and their caricatures usually include eyeglasses. Similarly, the Air Force is so associated with good vision that many would-be volunteers wrongfully self-select themselves out of the Air Force volunteer population because of their myopia. This only serves to exaggerate the difference. In other words, one would expect the recruits to have "better" vision than the students.
>
> A more subtle sampling difference, however, is seen when an ostensibly objective screening criterion is applied. If only those students and recruits are chosen whose far acuity is 20/25 or better the statistics are as shown in Table [2]. The most interesting difference is that of the far points. Because the means of the recruits' and students' far point distributions are separated by 1.5 D, limiting both distributions at one fairly extreme point (20/25 acuity) produces two new distributions with means still 0.3 D apart. This, in turn, leads to mean measured dark focuses also separated by approximately 0.3 D as shown.

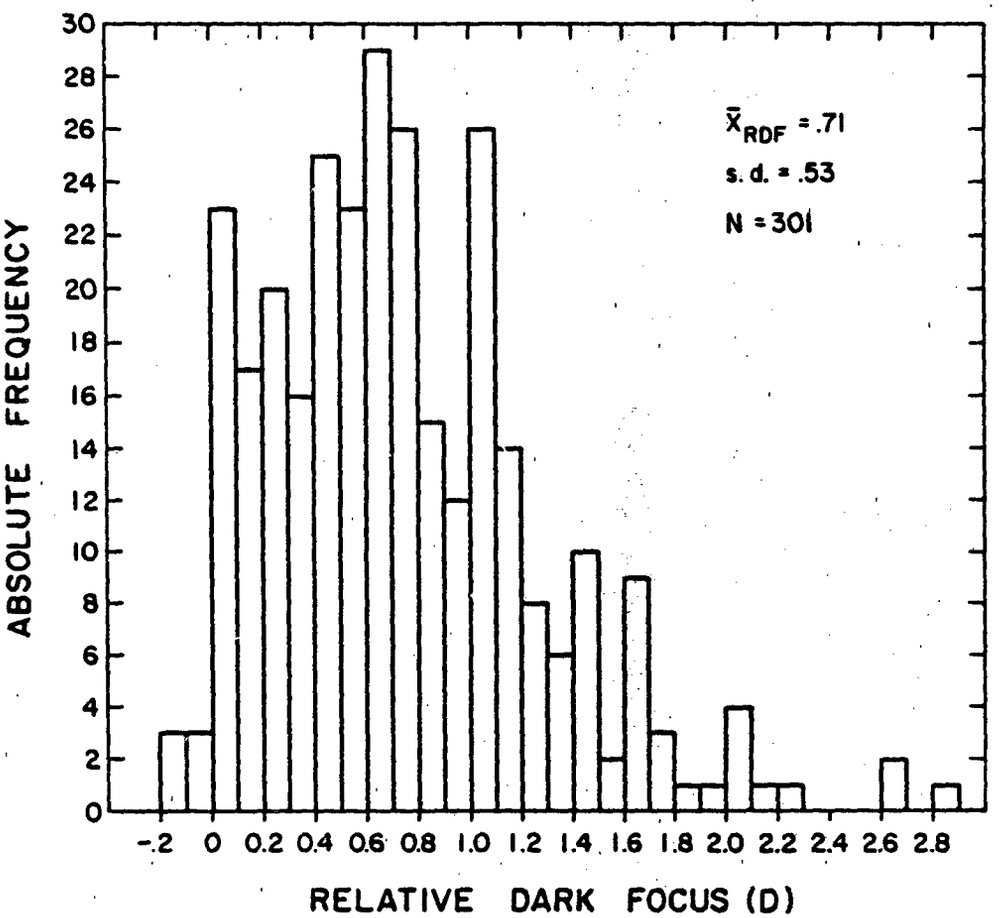


Figure 13. Distribution of relative dark focuses (Simonelli, 1979b).

Table 1

Comparison of Visual Characteristics of Psychology Students and Air Force Recruits (Simonelli 1979b)

Characteristic	N	Mean	sd	t	df	p
Far Point						
Students	114	1.918	2.39	6.27	266	.000
Recruits	154	0.471	1.37			
Measured Dark Focus						
Students	114	2.672	2.57	5.92	266	.000
Recruits	154	1.191	1.50			
Relative Dark Focus						
Students	114	.753	.56	.51	266	.612
Recruits	154	.720	.51			
Near Point						
Students	114	11.226	3.70	1.21	255	.266
Recruits	143	10.706	3.17			
Amplitude						
Students	114	9.308	3.39	-2.38	255	.018
Recruits	143	10.262	3.02			

Table 2

Comparison of Visual Characteristics of Psychology Students and Air Force Recruits After Selecting Individuals with Far Acuity of 20/25 or Better (Simonelli 1979b)

Characteristic	N	Mean	sd	t	df	p
Far Point						
Students	34	.147	.48	1.78	136	.077
Recruits	104	-0.116	.82			
Measured Dark Focus						
Students	34	.853	.54	1.55	136	.123
Recruits	104	.592	.93			
Relative Dark Focus						
Students	34	.706	.47	-.02	136	.986
Recruits	104	.708	.52			
Near Point						
Students	34	8.600	2.65	-2.62	125	.010
Recruits	93	10.179	2.62			
Amplitude						
Students	34	8.653	2.69	-3.31	125	.001
Recruits	93	10.405	2.62			

> ....Only 34 "natural" emmetropes (defined for purposes here as 20/25 or better without corrective lenses) were found among the 114 psychology students. Most likely, a proportionately small number of natural emmetropes were among the hundreds of "functional" emmetropes [myopes wearing corrections] in the Leibowitz and Owens studies, but their data cannot be separately identified. (Simonelli, 1979b, pp. 110, 112, 115)

The Effects of Age

The effects of age on the various accommodative measures are shown in an overly simplified summary form in Figure 14. It has long been known that the near point recedes with age (and the "shortening" of the arms), but the slower recession of the dark focus and far point, while suspected, has not been accurately documented previously. Although Figure 14 represents all these changes as linear functions of age, Figure 15 shows what all airline pilots have learned to expect, namely, that the onset of the outward migration of the focus for any individual tends to be noticed rather suddenly during the mid 40s.

Acuity Demands

There is an abundance of credible experimental data suggesting that the accommodative response is an antagonistic compromise between the pull of the stimulus and the pull of the dark focus. Many have speculated and some have reported anecdotal evidence that the resolution of the compromise for the individual eyeball depends largely on the fineness of the discrimination required. For example, Brian Brown of the Smith-Kettlewell Institute of Visual Sciences noted informally that tracking recordings from the Crane-Cornsweet infrared optometer shifted outward in stepwise fashion as subjects attempted to read successive lines of a Snellen eye chart across the room (personal communication).

Another example, once more in Simonelli's words:

> An intriguing finding by Iavecchia et al. (1978) was the subjects' differential accommodation to the various outdoor scenes--all at essentially 0 D. The scenes mathematically varied from 0.03-0.00 D, but this minuscule difference should not account for the gross variation observed. Was the eye actually responding to the minute changes in the dioptric distances of the scenes? Or was there some compositional aspect of the views (resulting from the masking) that elicited different levels of accommodation? That is, when nearby, larger objects are prominent, perhaps their more easily recognized details (subtending larger visual angles) identify the objects in sufficient detail so that it is not necessary to force accommodation out to 0 D. The eye may be "lazy," as it has been referred to by some.

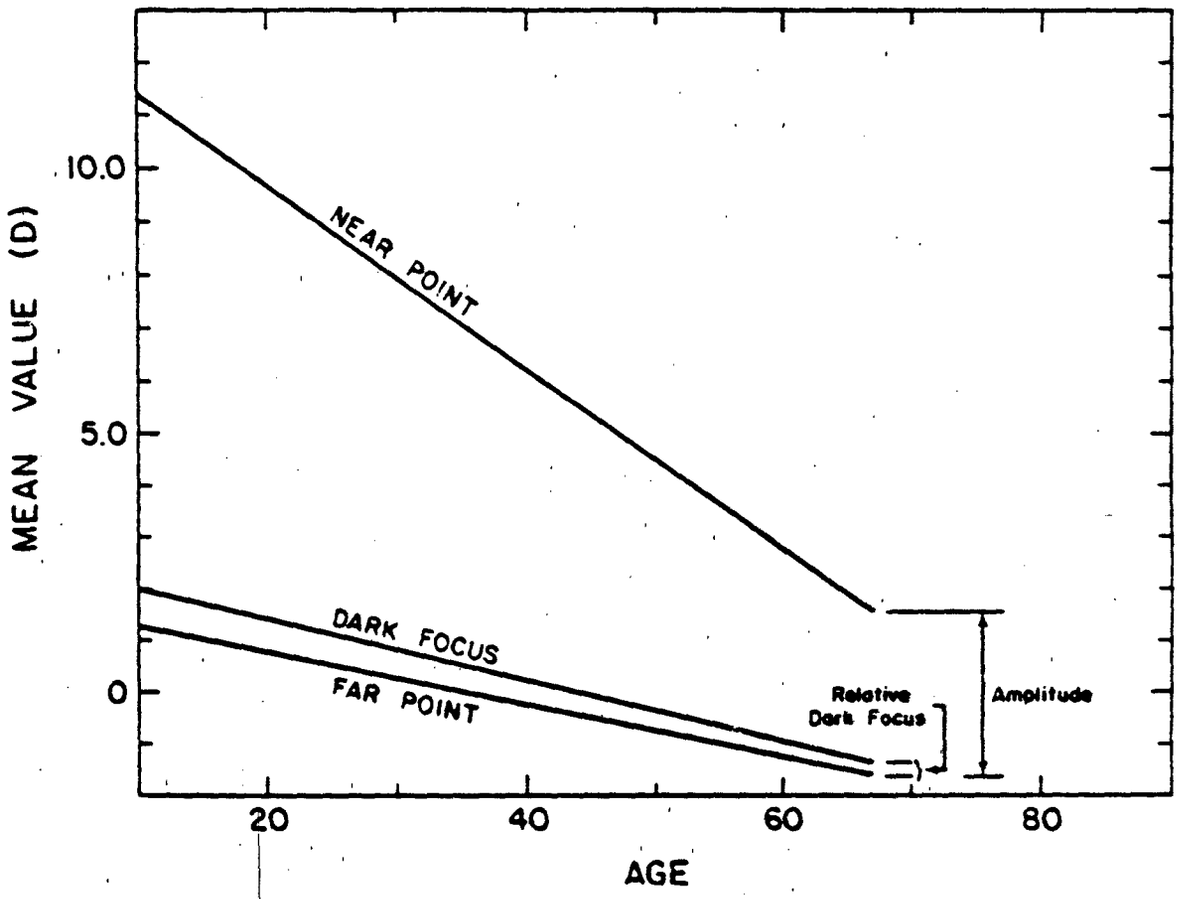


Figure 14. Basic changes in near point, far point, and dark focus with increasing age (Simonelli 1979b).

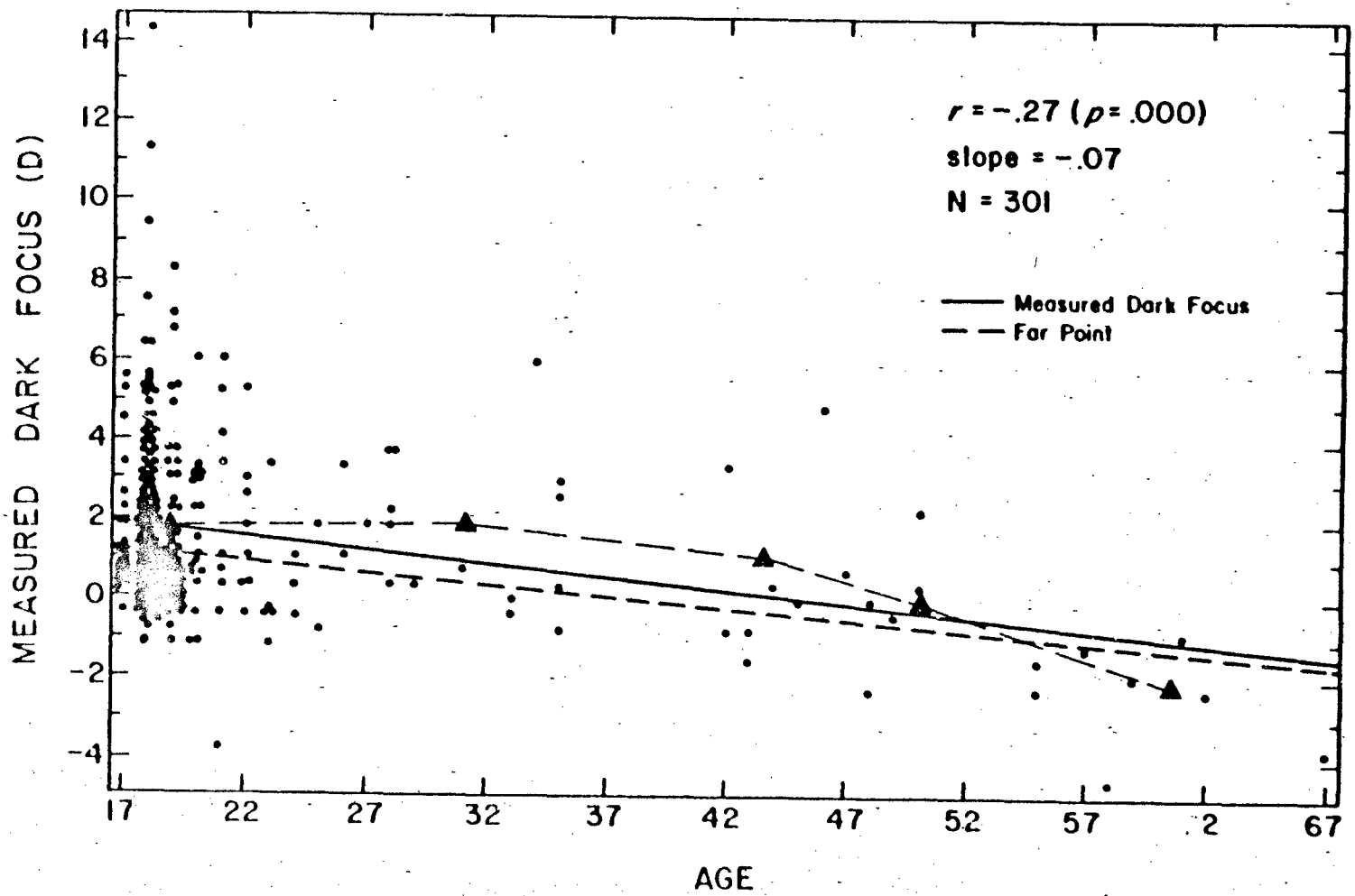


Figure 15. Relationship between measured dark focus and age showing the relatively rapid outward migration of accommodative responses during the mid 40s and thereafter. The linear regression lines for measured dark focus and far point recession with age do not accurately reflect the abruptness of the typical change for the individual.

> A response of 0.5 D to the roof of a large building 40 m away, as found in the Iavecchia study, may depend primarily on the "acuity demands" of the situation, the object of this experiment. Simply, the acuity demand of a target refers to the smallest details that must be resolved to recognize the target. Looking out an aircraft window at the blue sky does not pose much of a focusing challenge. Reading small print at a distance, however, requires more accurate focusing. What are the effects of such demands on accommodation and do these effects vary with the dark focus? (1979b, p. 72)

In Simonelli's experiment, acuity demands for accurate discrimination were varied while target distance (7.6 m) and illumination (1.3 FtL) were held constant. Targets were of two types, Snellen letters and modified Landolt Cs, as illustrated in Figure 16. Because Snellen letters vary in size as well as spatial frequency components, Landolt Cs of constant size but varying gap widths were included as an experimental control; if the Snellen letters elicited differing accommodative responses but the Cs did not, then stimulus size rather than acuity demand would appear to be the critical variable. This was not the case, and the results for the Snellen letters only are shown in Figure 17.

There were wide individual differences in the responses of the subjects, as illustrated dramatically in Figure 18, but on average accommodation shifted outward in an orderly stepwise progression until individual eyes could no longer resolve the letters and resist the inward pull of the dark focus. The confirmation of Brown's informal finding was not surprising, but the absolute values of the subjects' responses were: only the largest letters failed to pull accommodation to the distance of the target; as letters became more difficult to resolve, accommodation receded beyond the target distance. Presumably focusing at a distance greater than the target's would, if anything, reduce the clarity of the retinal image, while in fact acuity increased.

Zoom-Lens Hypothesis

Given the coincident facts that (1) accommodation shifts outward with increasing acuity demand, until resolution is no longer possible, and (2) the apparent size of targets increases disproportionately with outward accommodation, it is increasingly tempting to advance the hypothesis that outward accommodation, beyond the distance at which everything comes into clear focus, is functionally analogous to the action of a zoom lens of a television camera. More specifically, the hypothesis is that outward accommodation beyond optical infinity magnifies the retinal image, thereby causing it to impinge on a larger area of receptors and yield both finer discrimination and larger than life-size appearance.

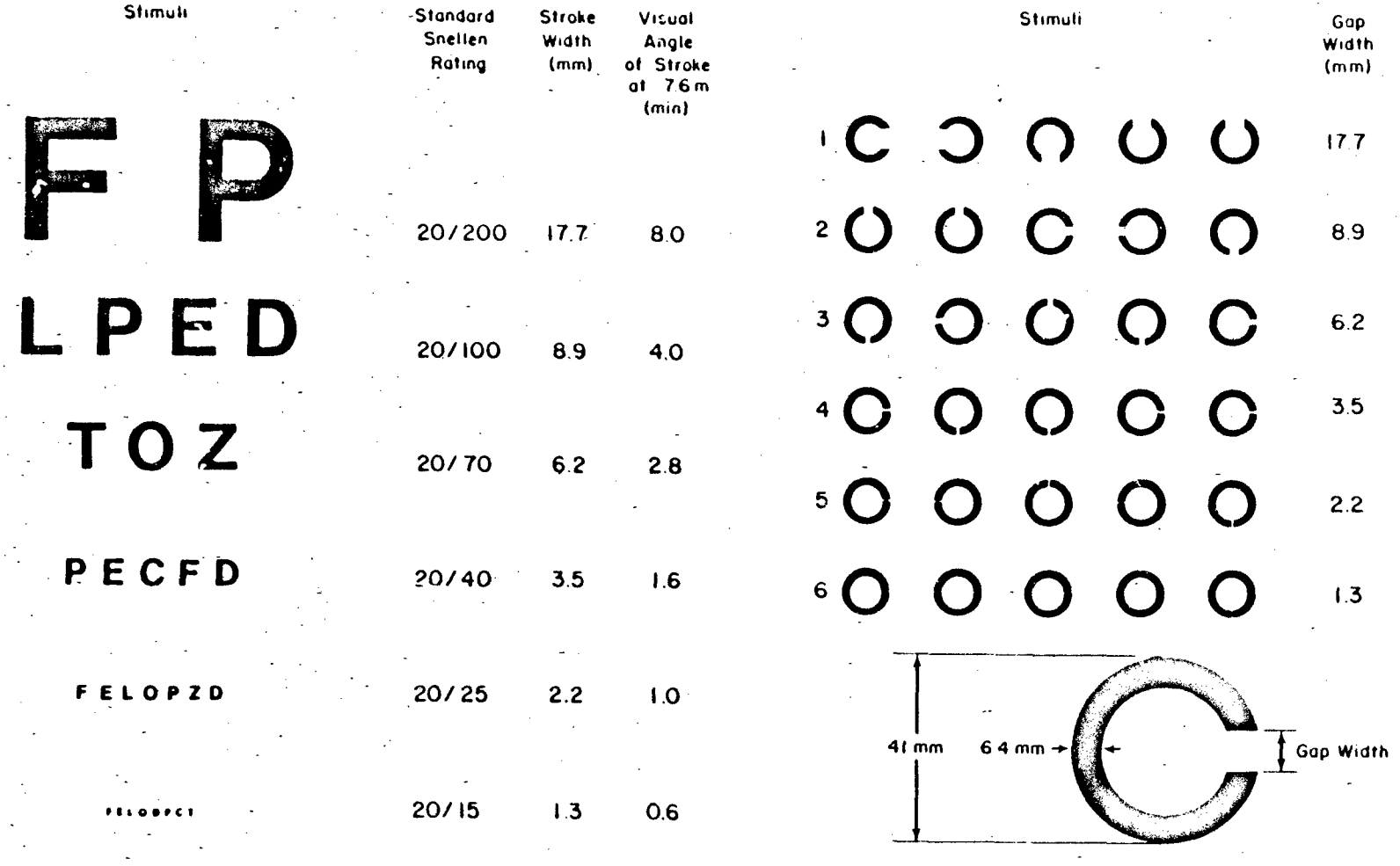


Figure 16. Snellen letters and modified Landolt Cs used by Simonelli to vary accommodative demand.

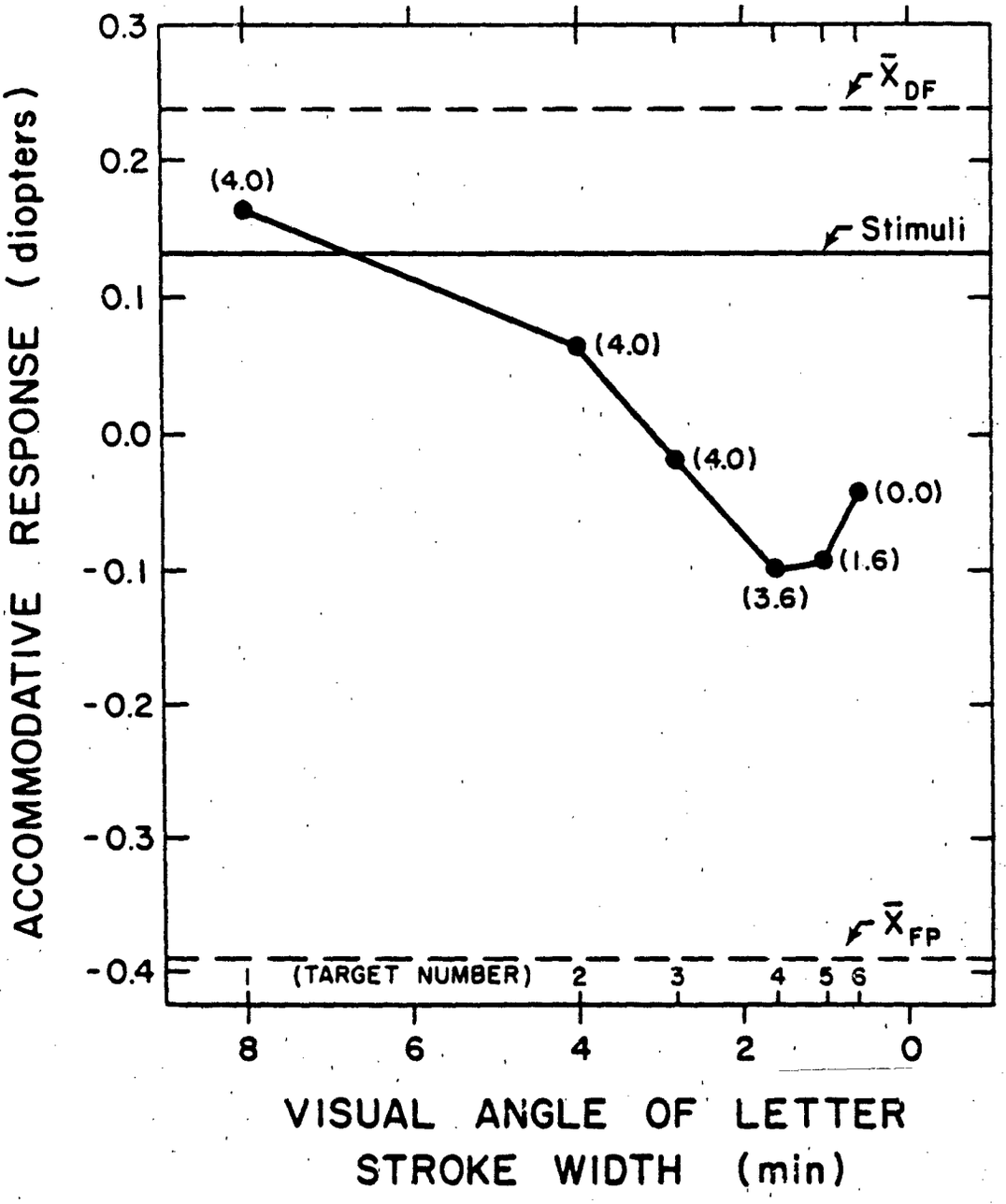


Figure 17. Mean responses to the Snellen targets. Numbers in parentheses are mean acuity scores, 4.0 representing all letters read correctly, 0.0 representing none read correctly.

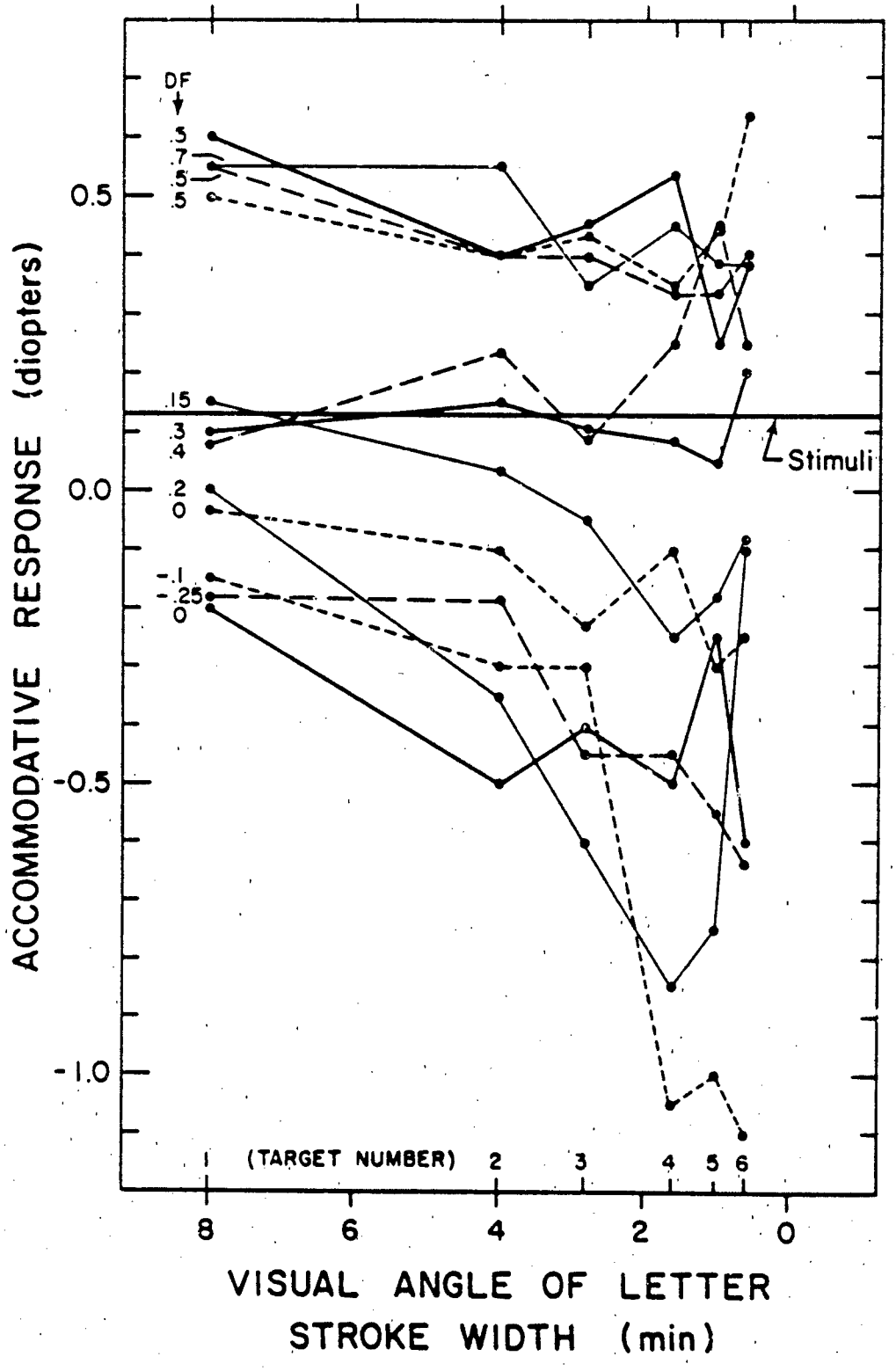


Figure 18. Individual accommodative responses to the Snellen targets showing the strong dependence of an individual's ability to accommodate beyond the stimulus distance on his or her resting accommodation distance, or dark focus (DF).

A critical test of the zoom-lens hypothesis is needed, but to date none has been reported. In my judgment the most nearly critical experiments are those dealing with the projection of afterimages, particularly those of Francis Young (1952) involving projection over large distances. Young's results would be predicted by the hypothesis, as would those of Ohwaki (1955). The assumption of invariant retinal image size with changing accommodation is essential to the validity of the model of the reduced schematic eye, but there is little support for this "null" hypothesis. To the contrary, the literature is replete with unexplained violations of size-distance invariance that would be predicted by the zoom-lens hypothesis.

The Critical Test

A critical test of a causal dependence of apparent size on accommodation distance would seem to require the independent manipulation of the latter without providing any other cues that might affect the former. A Pavlovian conditioning technique developed by Randle (1970) offers such a possibility. By using the output signal from the covert infrared tracking optometer to modulate the pitch of an audible tone, a subject can be provided nonvisual feedback of accommodative responses. As a subject learns the relationship between pitch and accommodation, the visual stimuli that initially induce accommodation can be eliminated in a classical conditioning arrangement.

A subject so conditioned accommodates to the tone with no visual cue to distance that might affect apparent size. By having subjects view the collimated virtual image of the moon machine against a dark background while accommodated to various distances, any changes in apparent size, relative to the adjustable real image of the comparison at a visible distance of one meter, would have to be caused by the aurally induced changes in accommodation. Alternatively, by a slight modification of the moon machine, afterimages can be formed while the eyes are accommodated to various distances and their sizes measured by a similar comparison procedure.

Accommodation, Personality, and Autonomic Balance

Robert Randle of Ames Research Center is among the most experienced hands-on investigators of visual accommodation and its correlates. It was he that first put me onto the importance of the autonomic nervous system in visual performance. We were studying a strip-chart recording from the Crane-Cornsweet infrared optometer when he casually commented, "This subject is obviously a sympathetic type." In the discussion that followed, Randle asserted that he could tell from a subject's behavior when he came into the lab and was introduced to the optometer what his accommodative responses would look like on the strip chart. In view of Randle's typical reluctance to speculate, his sympathetic-parasympathetic personality theory warranted more formal statement (Roscoe and Benel 1978) and systematic investigation (Gawron 1979).

Correlational Evidence

When Valerie Gawron accepted the challenge of this complex experimental problem, she soon found that Randle's informal observations had been partially anticipated by many and studied experimentally by a few, though apparently no one had actually reported correlating autonomic balance and accommodative responses. The prior work primarily involved attempts to relate indices of autonomic balance to various "personality" measures and clinically observed behavioral characteristics and, in a few cases, to intelligence and skilled performance. Indeed, Francis Young, another of the most experienced hands-on investigators of accommodation had specifically advanced the notion that intelligence and myopia are positively correlated (Young 1957).

The balance between sympathetic (SNS) and parasympathetic (PNS) branches of the autonomic nervous system has been linked to individual differences in personality by Wenger (1947; Wenger and Cullen 1972) and by Eysenck (1953), among many others. More specifically extraversion seems to be associated with SNS-dominance and introversion with PNS-dominance, which is basically what Randle had inferred. Since autonomic balance also mediates near (PNS) and far (SNS) accommodation (Cogan 1937; Olmsted 1944), the personality differences commonly attributed to near- and far-sighted people may depend upon a common underlying cause (Roscoe and Benel 1978).

Gawron's Findings

Gawron tested 152 recruits between the ages of 18 and 28 at Chanute Air Force Base during the summer of 1979. She investigated relationships among autonomic balance, refractive error, and introversion-extraversion as indicated by the Eysenck Personality Inventory. To measure autonomic balance she modified four physiological tests developed by Wenger and Ellington (1943) and processed the respiration and heart rate data to obtain an index of "weighted coherence" more recently introduced by Porges (1976). Her measures of refractive error included near point, using the RAF Near Point Rule, and dark focus and far point, using Simonelli's polarized vernier optometer.

What Gawron found was that some postulated relationships do indeed exist, as shown in Table 3, but they account for a relatively small fraction of the total response variance (except for the obviously strong interdependence of dark focus and far point). Specifically, the relationship between autonomic balance and near and far sightedness was reliably supported in the case of Porges's analytically sophisticated weighted coherence index (C_w) but was reliably contradicted by Wenger's index of autonomic balance ($\overline{A}$). This finding is puzzling in the view of the reliable correlation between $\overline{A}$ and C_w, as expected. Evidently some factor not measured or even identified is also at work.

Table 3

Correlation Matrix of Baseline Physiological and Personality Measures

	$\bar{A}$	C_w	I/E	DF	DFR	NP
$\bar{A}$ (Wenger's index)						
C_w (Porges's index)	.20*					
I/E (Eysenck)	.10	.05				
Mean Dark Focus (DF)	-.21	-.15	.03			
Dark-Focus Range (DFR)	-.01	-.06	-.16*	.07		
Near Point (NP)	-.04	-.18*	.00	.26**	-.10	
Far Point (FP)	-.18*	-.12	.06	.94**	-.03	.29**

*$p < .05$; **$p < .01$

Many investigators have noted apparently systematic relationships between autonomic and behavioral responses. Wenger (1947) emphasized the bidirectionality of such effects: autonomic responses affect behavior, and conversely behavioral events trigger autonomic responses. Porges (in press) speculates that C_w (his "weighted coherence" measure) is an index of central cognitive processing ability, and $\hat{V}$ (his measure of "vagal tone") reflects peripheral autonomic activity. Malmstrom (1978) has shown that performances of a simple information processing task causes a systematic outward shift in visual accommodation, an SNS response. The issues are: (1) Is information processing ability positively related to PNS dominance? and (2) Does task performance elicit an SNS shift in autonomic response?

Gawron addressed these issues by comparing the pre- and post-test autonomic responses of an experimental group, who performed a short-term memory task (delayed digit cancelling) for 4 minutes, with the pre- and post-rest autonomic responses of a control group. The majority of shifts in autonomic responses, including the relative dark focus as shown in Figure 19, were in the SNS direction following task performance; pre- and post-rest measures varied unsystematically in most cases. Incorrect response latencies and latency variability were reliably correlated in the predicted direction with baseline (pre-task) measures on both of Porges's indices (C_w and $\hat{V}$) but not on Wenger's index ($\bar{A}$).

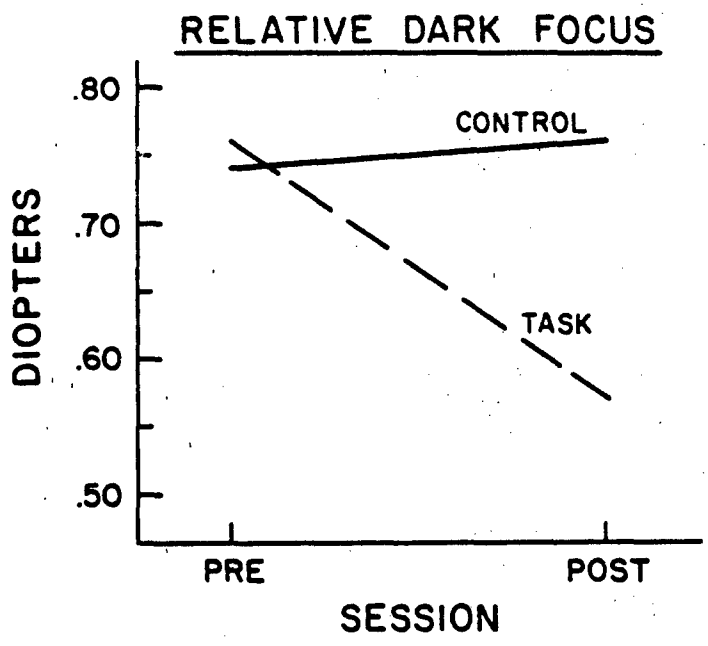


Figure 19. Relative dark focus in diopters before and after task performance (for task subjects) or rest (for control subjects).

In Gawron's words:

> Randle's hypothesis predicts that differential shifts in autonomic balance will be produced by the performance of cognitive tasks (SNS shift) versus rest and relaxation (PNS shift). Reliable SNS shifts did occur after task performance. This supports Randle's contention and replicates Malstrom's (1978) finding that task performance has an SNS effect. Application of these findings to the real world of aviation is as yet tentative but ...[the findings]... suggest the importance of future work in this area, particularly on the effects of elevated cockpit workload on pilots' approaches to landings.

FUTURE DIRECTIONS

This investigation has been concluded at the University of Illinois and has been resumed at New Mexico State University under a new grant from the Air Force Office of Scientific Research. The laboratory facilities developed under this program, including the moon machine and various optometers and cameras, were generously transferred to New Mexico State by the Department of Psychology of the University of Illinois. In addition, the Ames Research Center, NASA, has loaned New Mexico State the Crane-Cornsweet infrared tracking optometer used by Randle and his many associates. The exploratory investigations at Illinois were essential preliminaries to more systematic experiments in the continuing investigation of ground-referenced visual orientation.

ACKNOWLEDGEMENT

This research was performed at the University of Illinois at Urbana-Champaign under contract USAF F49620-77-C-0117 from the Air Force Office of Scientific Research with Professor Stanley N. Roscoe as principal investigator. Dr. Alfred R. Fregly was the scientific monitor for AFOSR, and Major Robert Eggleston and the late Dr. Sheldon MacLeod of the USAF Aerospace Medical Research Laboratory provided scientific guidance and contributions.

REFERENCES

Baird, J. C. Psychophysical analysis of visual space. Oxford: Pergamon Press, 1970.

Benel, R. A. Visual accommodation, the Mandelbaum effect, and apparent size (Tech. Rep. BEL-79-1/AFOSR-79-5). Las Cruces, NM: New Mexico State University, Behavioral Engineering Laboratory, 1979.

Benel, R. A., & Benel, D. C. R. Accommodation in untextured stimulus fields (Tech. Rep. Eng Psy-79-1/AFOSR-79-1). Champaign, IL: University of Illinois, Department of Psychology, 1979.

Cogan, D. G. Accommodation and the autonomic nervous system. Archives of Ophthalmology, 1937, 18, 739-766.

Cornsweet, T. N., & Crane, H. D. Servo-controlled infrared optometer. Journal of the Optical Society of America, 1970, 60, 548-554.

Duke-Elder, W. S. Textbook of ophthalmology Vol. 5. St. Louis: C. V. Mosby, Co., 1970.

Eysenck, H. J. The structure of human personality. London: Methuen, 1953.

Gawron, V. J. Eye accommodation, personality, and autonomic balance (Tech. Rep. BEL-79-2/AFOSR-79-6). Las Cruces, NM: New Mexico State University, Behavioral Engineering Laboratory, 1979.

Helmholtz, H. von Physiological optics. New York: Dover, 1867/1962.

Hull, J. C., Gill, R. T., & Roscoe, S. N. Locus of the stimulus to visual accommodation: Where in the world, or where in the eye? Tech. Rep. BEL-79-5/AFOSR-79-9). Las Cruces, NM: New Mexico State University, Behavioral Engineering Laboratory, 1979.

Iavecchia, J. H., Iavecchia, H. P., & Roscoe, S. N. The moon illusion: Apparent size and visual accommodation distance (Tech. Rep. Eng Psy-78-4/AFOSR-78-3). Champaign, IL: University of Illinois, Department of Psychology, 1978.

Kaufman, L. & Rock, I. The moon illusion. Scientific American, 1962, 207, 120-130.

Leibowitz, H. W., & Hennessy, R. T. The laser optometer and some implications for behavioral research. American Psychologist, 1975, 30, 349-352.

Leibowitz, H. W., & Owens, D. A. New evidence for the intermediate position of relaxed accommodation. Documenta Ophthalmologica, 1978, 46, 133-147.

Malmstrom, F. V. Effects of concurrent mental activity on static and dynamic accommodation responses. Unpublished doctoral dissertation, Oklahoma State University, Stillwater, Oklahoma, 1978.

Mandelbaum, J. An accommodation phenomenon. Archives of Ophthalmology, 1960, 63, 923-926.

Moses, R. Vernier optometer. Journal of the Optical Society of America, 1971, 61, 1593.

Ogle, K. N. Binocular vision. Philadelphia: Saunders, 1950.

Ohwaki, S. On the factors determining accommodation: Research on size constancy phenomenon. Tohuku Psychological Folia, 1955, 14, 147-158.

Olmsted, J. M. D. The role of the autonomic nervous system in accommodation for near and far vision. Journal of Nervous and Mental Disease, 1944, 99, 794-798.

Owens, D. A. The Mandelbaum effect: Evidence for an accommodation bias toward intermediate viewing distances. Journal of the Optical Society of America, 1979, 69, 646-652.

Palmer, E., & Cronn, E. Touchdown performance with a computer graphics night visual attachment. Proceedings of the AIAA Visual and Motion Simulation Conference. New York: Institute of Aeronautics and Astronautics, Paper 73-927, 1973.

Porges, S. W. Peripheral and neurochemical parallels of psychopathy: A psychophysiological model relating autonomic imbalance to hyperactivity, psychopathy, and autism. Advances in Child Development and Behavior, 1976, 11, 35-65.

Randle, R. J. Volitional control of visual accommodation. AGARD Proceedings No. 82 on Adaptation and Acclimatization in Aerospace Medicine. Neuilly-sur-Seine, France: North Atlantic Treaty Organization, 1970.

Randle, R. J., Roscoe, S. N., & Petitt, J. Effects of accommodation and modification on aimpoint estimation in a simulated landing task. Washington, D.C.: National Aeronautics and Space Administration, NASA Technical Note, in press.

Roscoe, S. N. How big the moon, how fat the eye? (Tech. Rep. ARL-77-2/AFOSR-77-2). Champaign, IL: University of Illinois, Aviation Research Laboratory, 1977.

Roscoe, S. N. When day is done and shadows fall, we miss the airport most of all. Human Factors, 1979, 21, 721-731.

Roscoe, S. N., & Benel, R. A. Is the eye smart or the brain forgiving? (Tech. Rep. Eng Psy-78-1/AFOSR-78-1). Champaign, IL: University of Illinois, Department of Psychology, 1978.

Roscoe, S. N., Hasler, S. G., & Dougherty, D. J. Flight by periscope: Making takeoffs and landings; the influence of image magnification, practice, and various conditions of flight. Human Factors, 1966, 8, 13-40.

Roscoe, S. N., Olzak, L. A., & Randle, R. J. Ground-referenced visual orientation with imaging displays: Monocular versus binocular accommodation and judgments of relative size. Proceedings of the AGARD Aero-Space Medical Panel Specialists Meeting, Athens, Greece. Neuilly-sur-Seine, France: North Atlantic Treaty Organization, 1976.

Simonelli, N. M. Polarized vernier optometer (Tech. Rep. BEL-79-4/AFOSR-79-8). Las Cruces, NM: New Mexico State University, Behavioral Engineering Laboratory, 1979a.

Simonelli, N. M. The dark focus of visual accommodation: Its existence, its measurement, its effects (Tech. Rep. BEL-79-3/AFOSR-79-7). Las Cruces, NM: New Mexico State University, Behavioral Engineering Laboratory, 1979b.

Simonelli, N. M., & Roscoe, S. N. Apparent size and visual accommodation under day and night conditions (Tech. Rep. Eng Psy-79-3/AFOSR-79-3). Champaign, IL: University of Illinois, Department of Psychology, 1979.

Weintraub, D. J., & Gardner, G. T. Emmert's laws: Size constancy versus optical geometry. American Journal of Psychology, 1970, 83, 40-54.

Wenger, M. A. Preliminary study of the significance of measures of autonomic balance. Psychosomatic Medicine, 1947, 9, 301-309.

Wenger, M. A., & Cullen, T. D. Studies of autonomic balance in children and adults. In N. S. Greenfield and R. A. Sternbach (Eds.), Handbook of psychophysiology. New York: Holt, Rinehart, and Winston, 1972.

Wenger, M. A., & Ellington, M. The measurement of autonomic and normative data. Psychosomatic Medicine, 1943, 5, 241-253.

Young, F. A. Studies of the projected after-images: III. The projection over large distances. Journal of General Psychology, 1952, 47, 207-212.

Young, F. A. Myopia and personality. American Journal of Optometry, 1957, 47, 714-718.

Multidimensional Scaling Analysis of Unit Performance in Field Exercises

Steven M. Medlin
Army Research Institute for Behavioral and Social Science
Alexandria, VA 22333

Paul Thompson
Department of Psychology
University of North Carolina
Chapel Hill, N.C. 27514

Abstract

The evaluation of unit performance in field exercises is based on the Training and Evaluation Outlines (T&EO's) contained in the Army Training and Evaluation Program manual. The contents of the manual, however, were selected by military judges without using standardized or scientific procedures for determining the tasks, subtasks, and standards. The present research applied statistical data analysis techniques to expert judgements in order to explore systematic methods for incorporating military expert opinion into unit evaluation procedures. The results indicated that judges used only three dimensions to evaluate unit performance, and that the dominant dimension was quality of overall performance. Several interpretations of the results were presented.

For several years the Army Research Institute for the Behavioral and Social Sciences has been involved in a systematic program of research on unit evaluation. The goal of the research is to develop a criterion-referenced system for evaluation of unit tactical performance. The system must be consistent with and be built upon the framework provided by the Army Training and Evaluation Program (ARTEP), because the ARTEP provides an essential link to doctrine and mission requirements. The system must go beyond current ARTEP methodology, however, incorporating its strengths and correcting its deficiencies.

As currently used, the ARTEP includes an ARTEP manual, a field exercise evaluation, and a training program based on the evaluation. The ARTEP manual drives the evaluation system. Traditionally, the ARTEP manual and the front-end analysis of behavioral objectives, performance variables, and measurements have relied heavily on expert military judgements. The current content of the Training and Evaluation Outlines (T&EOs) was selected by military experts without the benefit of available procedures for insuring concensus among different teams of experts.

ARI is engaged in a program of research to study the use of standardized, scientific procedures based on empirical data as a way to increase experts' concensus, as a means of testing and verifying expert opinion, and as a basis for reducing expert opinions down to a manageable set of variables that are most important in assessing unit performance in field exercises.

Previous research (Wheaton, Fingerman, Boycan, 1978) has shown that not all variables related to unit performance need to be measured and evaluated in order to assess unit tactical performance in field exercises. Methods have been developed for a logical and/or empirical basis for

limiting the number of such variables. However, a common problem with the methods of the above studies is that military experts must intially consider and list all the behaviors that could possibly influence unit performance. This method of generating performance constructs and measures has several disadvantages. First, the domain of constructs and measures often is quite large and unmanageable. Second, the behaviors listed are those that military experts suggest might possibly influence unit performance, but are they the factors or dimensions that evaluators actually use to assess unit performance?

The studies reported in this paper used methods which are designed to eliminate the need for full, exhaustive listing of training objectives. The methods are also designed to identify those dimensions or variables which evaluators actually use.

The research involved two related studies designed to define the major dimensions, or factors, that military judges use to assess unit performance in field exercises. In the first study, military judges (N=15) rated unit performance as described in written narratives of fifteen field exercises. These ratings were then analyzed using multi-dimensional scaling (MDS) techniques to determine how many dimensions the judges used to evaluate performance. The second study, attempted to define, or label, the dimensions obtained in the first study. To do this a list of possible dimensions defining attributes (adjectives or descriptive phrases) was composed, a new group of military judges (N=30) ranked the fifteen narratives with respect to how much each narrative was characterized by the attributes, and these ratings were used as input to a second MDS analysis.

Method

Experiment 1

Subjects. Fifteen Army officers, ranking from 1st Lieutenant to Major, served as subjects. All of the officers had served in the Army for at least five years, were college graduates, and had experience participating in and evaluating field exercises.

Procedure. Using written narratives of fifteen different armor/anti-armor field exercises, military judges rated each unit's performance. After an intial four-hour introductory session in which the task was explained to each judge individually and all questions were answered as clearly and completely as possible, each judge rated the narratives at his own pace over a two week time period. At the end of two weeks, the participants returned their ratings and were debriefed fully. To provide initial similarity judgements for the MDS analysis, each one of the narratives was used as a "target", and the other narratives were compared to the target along certain specific criteria. For instance, the officers were asked to "judge how similar each narrative is to the target narrative with respect to the performance of the combat unit in the narratives." Each of the fifteen narratives was used as the target; comparisons were made between all the other narratives and the target. A matrix composed of these ratings, called a similarity matrix, was used as input to the MDS computer program, ALSCAL (Takane, Young, and de Leeuw, 1977).

Results. Solutions in 1 to 6 dimensions were obtained and examined. The decision about the dimensionality, i.e., the number of dimensions, or

factors, of the solution was made after examining stress values and a correlation analysis indicating when an additional dimension was already accounted for by previously included dimensions.

An analysis of stress values was not conclusive in this case. When the stress values from all six solutions were plotted, there was no "break" or "bend" in the curve to indicate the dimensional value required to adequately represent the data (Kruskal and Wish, 1978).

In the canonical correlation analysis, the comparisons of the one- and two-dimensional solutions and the two- and three-dimensional solutions indicated that at least three dimensions were required for accurate representation of the data in a space. The comparison of the three- and four-dimensional solutions was not so straightforward. Dimensions 1 and 2 were predicted well from the first two canonical variates, but dimensions 3 and 4 were predicted moderately well from the third canonical variate, with the third dimension being better predicted than the fourth. This result suggested that in going from a three- to a four-dimensional space, the third dimension separated into two dimensions, and that some of the fourth dimension was already accounted for in the three-dimensional solution. Thus, the canonical correlation analysis suggested that a three-dimensional space is appropriate for the configuration of points representing the narratives.

The MDS analysis of narrative rating data indicated that military judges used three dimensions, or factors, to assess unit performance. The configuration of points in the space represents the relationships between the exercise narratives and the dimensions used to evaluate the narratives. After a configuration is obtained, the interpretation process begins. Two interpretation procedures are cluster analysis and the insertion of vectors corresponding to descriptive phrases into the space. To perform these analyses, however, a second study designed to collect additional data was necessary.

Experiment 2

Subjects. Thirty Army officers, ranking from 1st Lieutenant to Major, served as subjects. The officers participated in the study as part of their course work at the Armor Officers' Advanced Course at Ft. Knox, Kentucky. All of the officers had served in the Army for at least five years, were college graduates, and had experience participating in and evaluating field exercises.

Procedure. As part of Experiment 1, the participants were asked to list several aspects of unit performance that they considered in rating the narratives with respect to the target narrative. The twelve most frequently cited phrases were then used as target adjectives or descriptions in Experiment 2. During an initial one hour introductory session, the task was explained to each judge individually and all questions were answered as clearly and completely as possible. Each judge then rated the narratives at his own pace over a two week time period. At the end of two weeks, the participants returned their ratings and were debriefed fully. The thirty participating officers were asked to rank each narrative according to how well the unit performed each target phrase. For instance, the officers were asked to "rate these narratives on how well the unit conducts movement/uses covered and concealed routes". Each officer ranked each of

the fifteen narratives according to how well the combat units performed each of twelve descriptive phrases, or aspects of unit performance in field exercises. Data from these rankings were used in the following analyses designed to aid in the interpretation of the configuration obtained in Experiment 1.

Cluster analysis results. A cluster analysis of the points in the three-dimensional space resulted in five clusters. These groups seemed to be differentiated primarily on the basis of the quality of the unit's performance. The mean ratings for each narrative on each adjective were calculated. Group 1 narratives were rated highly on all twelve adjectives; Group 2, 3, and 4 narratives were in the middle ranks on all adjectives; and Group 5 narratives were rated very low on all adjectives. The variation found in the middle three groups was quite high, but their rankings were almost always below Group 1 and above Group 5.

The descriptive phrases were cited by participants in Experiment 1 as being behaviors that were salient or important in evaluating unit performance in the field exercise narratives. Although twelve adjectives were mentioned frequently, the military judges seemed to use only one major dimension to assess performance - the quality of overall performance. The other two dimensions in the space were less variable, i.e., the points were not spread out on the dimension, and thus had to be of less importance. The cluster analysis suggested five groups of narratives that varied in quality of overall performance, but did not aid in identifying the two other dimensions in the space. A second interpretative analysis was needed to label these factors.

External unfolding procedures. A second procedure for interpreting a configuration is to insert vectors corresponding to the adjectives or descriptive phrases into the solution space. There are several advantages to using this method. The method is objective, as the data used to place the vector in the space are provided by subjects rather than the experimenter. The vectors are not confined to the axis which the MDS algorithm has selected, but may have any orientation in the space. The possible explanations are suggested by the persons who generate the data for the space, so their perceptions are used to explain their space. The process used to incorporate these adjectives or descriptive phrases into the configuration is known as external unfolding (Coombs, 1964).

External unfolding results. The external unfolding analysis (Medlin, and Thompson, in press) indicated that the three dimensions were: overall quality of performance, use of indirect fire, and the use of TOW's. The highly positively rated narratives and highly negatively rated narratives tended to dominate the solution, however. Thus, these points were excluded from one analysis to examine their effect on the unfolding solution. The strong quality of performance dimension was still present in this reduced set of narratives, but the best-fitting adjectives fell into two distinct groups, one composed of leadership functions and the other composed of tactical skills or functions.

Discussion

The studies present a complicated and somewhat confused picture of evaluation in field exercises. Several conclusions can be made, however. First, military judges get a general impression of unit performance, good,

bad, or average, and evaluate the units accordingly. Second, these general impressions are not based primarily on specific aspects of performance. Third, when no strong good or bad impressions are made, the judges differentiate performance on other dimensions.

Several interpretations of these results are possible. The results may be a methodological artifact due to insufficient information in the narratives or due to the biases of the evaluators who originally wrote the narratives. Performance on all aspects of unit tactical skills may change in unison, thus the dominance of the quality of overall performance dimension. The judges may use only a generalized quality of performance dimension, because they do not know what other dimensions to consider, how to assess performance on other dimensions, or how to assimilate information from the other dimensions to arrive at a single evaluation of unit performance. Further research on evaluator judgement processes needs to be conducted, particularly research that can be tightly controlled to provide a rigorous investigation of the judgemental decision process.

References

Coombs, C. A theory of data. New York: Wiley, 1964.

Kruskal, J., and Wish, M. Multidimensional Scaling. Beverly Hills, CA: Sage Publications, 1978.

Takane, Y., Young, F.W., and de Leeuw, J. Nonmetric individual differences mutlidimensional scaling: An alternating least-squares method with optional scaling features. Psychometrika, 1977, 42, 7-67.

Wheaton, G.R., Fingerman, P.W., and Boycan, G.G. Development of a Model Tank Gunnery Test. Army Research Institute Technical Report. (TR-78-A24), August, 1979.

Relationship of Supervisory and Job Variables to Organizational Effectiveness Variables

William H. Hendrix
Air Force Institute of Technology
Wright-Patterson AFB OH 45433

Abstract

Research focused on establishing relationships between managerial, situational variables and effectiveness variables. Seven significant ($p < .001$) relationships were identified.

Contingency approaches to organizational effectiveness generally consider effectiveness to be a function of the criteria of effectiveness, the situation, and the manager's style. Within this framework, no one criterion of effectiveness is postulate; rather many criteria may be appropriate depending on the other components of the model (i.e., the situation and the manager). Many researchers have noted that organizations operate within this type of model. Organizations have a variety and often contradictory goals (Dubin, 1976; Cameron, 1978). Effectiveness criteria at one organizational level may differ from those at other levels (Price, 1972), and criteria appropriate at one point in time may change and be less appropriate at later times.

This paper focuses first on research to identify those relationships which exist within the Air Force between managerial and situational variables, and variables associated with three criteria of effectiveness; namely, satisfaction, organizational climate, and perceived productivity. Once established these patterns could serve as a basis for developing hypotheses for further exploration of the dynamics associated with the relationships between the manager's style, the situational environment, and different criteria of effectiveness. In addition, research was performed to establish whether job related or managerial/supervisory variables were more predictive of the three criteria of effectiveness.

Method

Subjects

Subjects consisted of 4786 military and civilian personnel located at 5 Air Force bases representing six major commands. The sample's composition was: 86% males, 14% females; 17% officers, 66% enlisted, and 17% civilians.

Survey Instrument

Data were collected using the Organizational Assessment Package (OAP), an attitudinal survey primarily containing 7-point likert scales (Hendrix and Halverson, 1979). The OAP included measures of areas relating to the job, one's supervisor, the organizational climate, the perceived productivity of one's work group, and satisfaction.

Canonical Analysis

Canonical analysis was performed to establish relationships between criterion variables (Set A) and managerial and situational variables (Set B). The criterion variables were obtained from the OAP's Job Satisfaction Questionnaire, Organizational Climate Inventory, and Perceived Productivity

PRECEDING PAGE BLANK-NOT FILMED

Inventory. The managerial variables were from the OAP's Supervisor Inventory while the situational variables which dealt primarily with the job were obtained from the Job Inventory. In a previous study (Hendrix and Halverson, 1979), 22 OAP factors were extracted. For this study, the 50 top-loading criterion items (Set A) and the 50 top-loading managerial and situational items (Set B) for those 22 factors were used to establish relationships between the two sets of items

Regression Analysis

Factor scores were generated for three criteria of effectiveness (i.e., job satisfaction, organizational climate, and perceived productivity). In turn, managerial/supervisory variables and job-related variables were separately regressed against the three criteria to establish their predictive accuracy.

Results

Canonical Analysis

Seven multivariate relationships were found which accounted for a minimum of 10 percent of the total variance and were significant beyond the .0001 level using Bartlett's test.

Relationship 1. This relationship accounted for 73 percent of the data variance. This relationship was characterized as one where employees were satisfied with their job but perceived the climate to be poor in that the organization was not run as well as in the past and work planning was not effective. In this case, the supervisor did not stress accomplishing more than other work groups, but did expect his directions to be followed exactly, while representing the group at all times.

Relationship 2. Approximately 37 percent of the data variance was accounted for by this relationship. This relationship was one where the work group was frequently involved in crash projects and the work group members' perceived productivity was low in comparison to other work groups, and the organization was ineffective in planning work to be accomplished. In terms of job enrichment components, the job of work group members required use of a variety of skills and talents, and the job was perceived as significant in affecting others. The job, however, required the same tasks to be repeated within a short period of time.

Relationship 3. This relationship accounted for approximately 25 percent of the data variance. Characteristic of this relationship was a work group which had its complaints aired satisfactorily, but were not kept aware of important events and situations, and group members were not recognized for outstanding performance. In this case, individuals used a variety of skills and talents on their job, but it was not perceived as being significant in affecting others. Also, supervisors did not set high performance standards nor did the supervisor make his responsibilities to the group clear.

Relationship 4. Approximately 20 percent of the variance was accounted for by this relationship. This relationship was one where individuals were kept informed on important events and situations, but were not rewarded for outstanding performance, and were unsatisfied with opportunity to acquire valuable skills in their job which would prepare them for the future. The supervisor required that his directions be followed exactly, but the nature of the job was such that a great deal of freedom and independence in selecting procedures to accomplish it were permitted. The job required use

of management information systems such as computer printouts, had adequate information from other work groups for performing the job, and required little future planning of six months or more.

Relationship 5. This relationship accounted for approximately 19 percent of the data variance. For this relationship quantity of output by the work group was high, group members were proud to work for the organization, although the organization was perceived as being run better in the past, and group members were unsatisfied with information on policies such as promotions. The supervisor was perceived as one who requires directions to be followed exactly and overcontrols members' work. The job did not permit very much freedom in doing the work as the individual saw fit. Also, the job required little short-range planning (weekly or monthly).

Relationship 6. Approximately 15 percent of the data variance was accounted for by this relationship. This relationship involved those work groups which did not have complaints aired satisfactorily, were not recognized for outstanding performance, had good cooperation between work groups, and had members who were unsatisfied with their job security. In terms of the job and the supervisor, group members perceived the amount of work space provided to be inadequate, and that the supervisor overcontrolled their work, required his directions to be followed exactly, made his responsibilities clear, did not set high performance standards, and did not represent the group at all times.

Relationship 7. This relationship accounted for approximately 12 percent of the variance. On the criteria side, group members were satisfied with their job security and their work, perceived their productivity in terms of quality of output to be very high, and perceived the climate to be good in that the organization was very effective in its planning. The group's jobs required a large amount of time for weekly and monthly planning, yet had job interferrence due to additional duties. Group members felt they were being prepared for increased responsibilities and had supervisors whose directions had to be followed exactly.

Discussion. The seven patterns established through canonical analysis are at the very least difficult to interpret. The major relationship found when looking across the seven patterns was that job-related variables correlated more highly than the managerial/supervisory variables with the criterion variables of satisfaction and climate. Also, the perceived productivity criterion variables were highly related less frequently to the job-related and managerial/supervisory variables than were the satisfaction and climate variables. This can be noted in Table 1 which lists up to six of the top-loading variables which were greater than .70 for the criteria set (Set A) and the job-related, managerial/supervisory set (Set B). As difficult as the relationships are to interpret, the data tend to support a contingency approach where job-related variables appear to be more highly related to the criteria in some patterns, while in one pattern the managerial variables appear to be more dominant while still in others both take on nearly equal importance.

TABLE 1

Variable Frequency for Criteria, Job-Related, and Managerial/Supervisory Variables

Canonical Analysis Sets	Relationship 1	2	3	4	5	6	7	Total
Set A								
Criteria								
Satisfaction	3	3	2	3	2	3	2	18
Climate	2	1	4	3	3	3	2	18
Perceived Productivity	-	2	-	-	1	-	2	5
Set B								
Job Related	3	6	4	5	4	1	4	27
Managerial/Supervision	3	-	2	1	2	5	2	15

Regression Analysis

The fifty job-related variables from the Job Inventory and the forty-one supervisory variables from the Supervisor Inventory were regressed separately against the factor scores for the three criteria of job-related satisfaction, organizational climate, and perceived productivity. As can be noted in Table 2, job-related variables accounted for more of the criteria variance than did the managerial/supervisory variables.

TABLE 2

Job-Related and Managerial/Supervisory Variables Regressed Against Three Criteria[a]

Variables	CRITERIA Job-Related Satisfaction	Organizational Climate	Perceived Productivity
Job Related	.52	.52	.43
Managerial/ Supervisory	.27	.42	.30

[a]Cell entries are R^2's

Discussion. Regression analysis results tend to support the conclusion based on canonical analysis data that job-related variables were more highly related to the criteria of satisfaction and climate than were those dealing with management and supervision.

Summary

Seven relationship patterns were identified through canonical analysis between three criteria of organizational effectiveness and managerial/supervisory and job-related variables. The highest loading variables making

up these patterns differed for each pattern. For some patterns job variables were dominant in relating to criterion variables; while in one, managerial variables were dominant, and with others both played a significant role. Data from canonical analysis and regression analyses indicated that job-related variables to be in general more predictive of the criteria than were the managerial/supervisory variables.

References

Cameron, K. Measuring organizational effectiveness in institutions of higher learning. Administrative Science Quarterly, 1978, 23, 604-629.

Dubin, R. Organizational effectiveness: Some dilemmas of perspective. Organization and Administrative Sciences, 1976, 7, 7-14.

Hendrix, W. H. & Halverson, V. B. Organizational survey assessment package for Air Force organizations. AFHRL-TR-78-93. Brooks AFB, TX: Occupation and Manpower Research Division, Air Force Human Resources Laboratory, February 1979.

Price, J. L. The study of organizational effectiveness. Sociological Quarterly, 13, 3-15.

The Impact of Organizational Conditions on Enlisted Attrition: Perceptions of Navy Human Resource Management Specialists

Gerry L. Wilcove and Ervin W. Curtis
Navy Personnel Research and Development Center
San Diego, California 92152

Abstract

Navy Human Resource Management Specialists have a wealth of experiential information regarding the impact of organizational conditions on the premature attrition of first-term enlistees. A specially constructed questionnaire with 45 multiple choice items was administered to 70 Specialists functioning in several regions of the U.S. Five major problems were identified: crisis management, lack of job recognition, bureaucratic demands, feelings of powerlessness, and supervisory non-support. Open-ended questions identified additional problems (e.g., in-port shipboard inspections), as well as suggesting possible solutions.

The goal of the Human Resource Management Support System (HRMSS) is to develop and utilize the human resource potential in the Navy to the fullest extent possible. HRM Specialists represent the point of contact between individual operational units and the network of concepts and procedures constituting the HRMSS. Extensive and varied information sources are available to Specialists on the quality of organizational life and its impact on enlisted attrition. They administer and analyze data from a comprehensive HRM Survey completed by all organizational levels. They conduct feedback sessions with the commanding officer, executive officer, departmental heads, and individual workgroups. Problems are explored in depth at such times, and Specialists later conduct command workshops designed to alleviate such problems.

The problem of concern here is that 32% of the enlisted personnel entering the Navy in FY76 attrited prematurely during the next three years. HRM Specialists, with their vast experience with command and lower level problems, should be able to provide valuable insights into the causes of attrition.

PROCEDURE

A questionnaire was constructed on the basis of the research literature and input from enlisted personnel. Seventy HRM Specialists functioning in several regions of the U.S. completed the questionnaire. Each of the Specialists had assisted 26 commands, on the average. On the questionnaire, Specialists rated the extent to which each of 45 conditions increase the desire of quality first-termers to attrite. A 5-point scale was provided, ranging from "a very little extent" (coded as 1) to "a very great extent" (coded as 5). The 45 conditions were grouped into six conceptual categories: (1) Family, (2) Navy System: Personal Consequences, (3) Navy System: Professional Consequences, (4) Navy System: "Total Life" Consequences, (5) Leadership, and (6) Job Satisfaction.

The opinions and assertions contained herein are those of the writers and are not to be construed as official or as reflecting the views of the Navy Department.

PRECEDING PAGE BLANK-NOT FILMED

Open-ended questions were also presented on the questionnaire to provide qualitative depth to the rating scale responses, to identify causes of attrition missing from the questionnaire, and to pinpoint other Navy problems of concern.

RESULTS

The results for the 45 conditions are presented in Table 1. The overall mean for each category or class is underlined. The average standard deviation for conditions within a category is also presented in the table and underlined.

No entire class of conditions was seen by the HRM Specialists as greatly contributing to the premature attrition of first-termers--all category means were smaller than 3.5. Leadership produced the largest mean--3.38, while "The Navy System: Total Life Consequences" yielded the second highest mean--3.32. This latter category represents those factors that affect both the personal and professional parts of an individual's life.

Although no class of conditions apparently contributes greatly to attrition, the Family was the sole one contributing only slightly according to the Specialists. Moreover, six specific Navy conditions yielded means over 3.5. Four were classified under Leadership, and included: (1) the disorganized nature of Navy procedures or of Navy leaders and supervisors (also classified under Navy System: Total Life Consequences), (2) lack of task support or emotional support from superiors, (3) lack of credit or recognition for a job well done, and (4) the extent to which "crisis management" techniques are employed. Specialists also perceived two Job Satisfaction conditions as contributing greatly to attrition: (1) feelings of powerlessness and (2) excessive time spent on bureaucratic demands.

The data revealed 12 conditions (identified with an L in the table) which apparently contribute only slightly to attrition (i.e., their means are less than 2.5). Three examples occur under Navy System: Professional Consequences. The first concerns educational opportunities; the second, organizational support for rating changes; and the third, the quality of technical training. A fourth example occurs under Leadership and concerns the supervisor's proficiency in enforcing work and other standards.

In the open-ended questions, HRM Specialists were asked to briefly discuss problems they thought contributed to attrition, problems which had been mentioned in the questionnaire, or others they deemed important. Twenty-one Specialists elaborated in some detail the need for improving the leadership skills of supervisors. The primary skill mentioned was treatment of subordinates. Nineteen Specialists mentioned in-port shipboard inspections. Although inspections frequently fulfill vital functions, they also are generally believed to constitute a bureaucratic demand, identified earlier in the paper as a major cause of attrition.

Seven Specialists (who had assisted over 200 commands) mentioned the motivation and assimilation of recruits as a problem. As one Specialist put it:

> If a new recruit is brought aboard as an important member of a team and given those things that everyone desires--companionship, respect as an individual, capable of contributing

Table 1

The Impact of Various Conditions and Categories of Conditions on Attrition as Perceived by HRM Specialists

Condition	Category	Mean[a,b]	S.D.
	Family	2.47	1.12
Pressure from family to leave the Navy		2.43 (L)	1.18
Family problems		2.51	1.06
	Navy System: Personal Consequences	2.56	1.07
Unsatisfactory food in the Navy		1.79(L)	.83
The lack of recreational opportunities		1.94(L)	.95
A dislike for regulations on personal grooming		2.03(L)	1.09
Inadequate medical care facilities and benefits		2.14(L)	1.09
Inadequate treatment of dependents or inadequate programs for dependents		2.46(L)	1.12
Poor living conditions in port		2.90	1.10
Poor living conditions at sea		2.99	1.17
Loss of freedom to conduct one's personal life		3.38	1.02
Separation from family		3.39	1.22
	Navy System: Professional Consequences	2.56	1.02
Limited educational opportunities in the Navy		2.13(L)	.98
The difficulty involved in changing ratings		2.37(L)	1.13
The low quality of, or the lack of opportunity for, technical training		2.38(L)	.87
Dissatisfaction with the Navy's performance evaluation process		2.50	1.06
Absence of a strong career counseling program		2.51	1.02
The trend towards lower standards of conduct and performance		2.57	1.05
Promotion opportunities limited or based on paygrade instead of ability and knowledge factors		2.76	1.08
Recruitment promises not kept		2.87	1.12
The assignment of jobs based on paygrade instead of ability and previous training or experience		2.99	.86
	Navy System: "Total Life" Consequences	3.32	1.13
Dislike for Navy rules and regulations		2.93	1.21
An excessive amount of sea duty/unsatisfactory rotation procedures		3.43	1.21
The disorganized nature of Navy procedures or of Navy leaders and supervisors		3.59(H)	.96
	Leadership	3.38	.97
Being falsely accused of irresponsibility, laziness, rule-breaking or other offenses		2.41(L)	1.15
Harassment by superiors (i.e., being verbally abused, "bullied", or "pushed around")		2.77	1.07
Unkept promises for additional training/education		2.99	1.01
Denial of personal, training, or job assignment requests that are especially important to the individual		3.18	.88
Lack of opportunity to discuss personal and work problems with superiors		3.18	1.05
Person does not see how his job fits into an overall plan or mission		3.46	.91
The disorganized nature of Navy procedures or of Navy leaders and supervisors		3.59(H)	.96
Lack of task support or emotional support from superiors		3.73(H)	.95
Lack of credit or recognition for a job well done		4.20(H)	.81
The extent to which "crisis management techniques are employed		4.26(H)	.86

Table 1 (continued)

Condition	Category	Mean[a,b]	S.D.
	Job Satisfaction	3.07	1.04
Quality of one's co-workers		2.31(L)	.93
Low Pay		2.32(L)	.95
The person feels that he has been required to work too many hours a week at sea		2.52	1.07
The perceived loss of fringe benefits		2.87	1.31
Poor working conditions in port		2.89	1.08
Poor working conditions at sea		2.89	1.06
Dissatisfaction at not being part of a closeknit, productive work team		3.08	1.11
The person feels that he has been required to work too many hours a week in port		3.30	.95
The lack of challenging work		3.32	1.08
Lack of personal fulfillment--for example, little opportunity to develop one's talents, to meet one's ideals, or to determine one's identity		3.36	1.12
A job assignment that is disliked		3.48	.88
Feelings of powerlessness--for example, the person is not permitted to function independently, cannot control the factors which determine productivity, cannot move from immediate work area, etc.		3.64(H)	.89
Excessive time spent on bureaucratic demands (inspections, paperwork, artificially-created tasks, etc.)		3.87(H)	1.05

[a]Specialists indicated for the Navy conditions listed in the questionnaire the extent to which each contributed to the desire of quality first-term personnel to leave the Navy early. Response options were coded as follows: 1 = to a very little extent, 2 = to a little extent, 3 = to some extent, 4 = to a great extent, 5 = to a very great extent.

[b]L = one of the lowest means (i.e., less than 2.5), H = one of the highest means (i.e., at least 3.5).

to a team effort, he/she can start as charged up as when leaving recruit training.

Other comments indicated that boredom, frustration, confusion, and the desire to attrite occur when recruits fail to receive the necessary attention.

Many other concerns were voiced besides supervisory behavior, inspections, and recruit motivation--but each Specialist seemed to focus on a different one (pay, planning deficiencies, communication, role definitions, etc.). Despite their diversity, however, such problems can be solved, according to the Specialists, if the Navy keeps in mind the following:

"Most personnel could live with most situations if their command and supervisors supported them fully."

"A very high percentage of quality men will work incredibly long and hard hours (for years on end), if they find high job satisfaction, see meaning in what they're doing and understand their

place in the overall scheme of things. Reasonable reward and recognition are also part of the grease that makes these gears mesh smoothly."

"The main reason [for attrition] seems [to be] a widespread perception . . . of a lack of suitable reward (in terms of achievement, growth, challenge, responsibility, advancement, etc.)."

The significance of these observations may extend beyond first-term attrition--14 Specialists identified the retention of quality career personnel as a major problem facing the Navy today.

Personal & Professional Identity

Frank R. Wood & David B. Porter
Department of Behavioral Sciences & Leadership
USAF Academy, Co 80840

Abstract

This paper proposes a model of professional identity development based on Erikson's model of personal identity development. Types of professional vitality and withdrawal are described and factors which promote these orientations are suggested.

Personal and professional identity are similar concepts and should follow a similar developmental process. This similarity was suggested by Super (1963) and again by Hall (1976) who compared the developmental stages of personal identity development with hypothetical stages of career development. While suggesting an interesting comparison, in our opinion these authors have only cursorily applied Erikson's ideas and by emphasizing the development aspects of a career, they have overlooked the non-linear qualities of "real" careers. As such, their efforts do little to explain phenomena such as commitment, attrition, retention, career change and productivity.

The central aspect of Erikson's (1969:96) model which is relevant to these problems is the concept of interactional crisis which is "not a threat of catastrophe, but a turning point, a crucial period of increased vulnerability...the ontogenetic source of generational strength or maladjustment." Depending on how these crises are resolved, the individual in a professional organization may either develop a sense of professional vitality or withdraw.

In Erikson's view, interactional conflict is brought about by conflict between individual needs and environmental constraints. It is the resolution of these conflicts that provides the impetus for change in the individual's orientation with the organization. Predicting the outcome of such conflict is not an easy task. Even in a small organization or group, each individual, because of his unique perceptions, attitudes and experiences, may perceive the environment significantly differently than his co-worker. In a large institution such as the Air Force, the variety of specific environments is practically infinite. Although the interactions of individuals with their specific environments are in many ways unique, they also share several common aspects.

In this paper, we propose to describe three ideal types of Air Force officers which are generated by individuals who have resolved their interactional crises in different ways. We will also suggest several organizational factors which influence the resolution of crises in the careers of Air Force officers in favor of professional vitality or withdrawal. Throughout, an attempt will be made to apply Erikson's concepts of personal identity development to the process of professional identity development in the Air Force. Finally, to demonstrate the value of this orientation, comments from recent surveys and interviews with Air Force junior officers will be examined for support of this theoretical orientation.

PRECEDING PAGE BLANK-NOT FILMED

A NEW APPROACH

Interactional crises can be resolved in favor of at least three alternatives: professional vitality, organizational withdrawal, or craft idiocy (see Diagram 1). While these alternatives represent ideal types of persons which may be found in the officer corps of the Air Force, they are not permanent states but rather reflect a continuum of alternatives which occur and reoccur throughout an individual's career. Each ideal type will be discussed and related to the environmental factors which promote crises resolution in its favor.

Professional vitality is characteristic of an individual who actively masters the environment by being both productive and innovative. His congruence reflects the successful integration of the social, psychological and occupational aspects of his life and provides him with resiliance and a high tolerance of ambiguity. His ability to perceive himself and the world correctly leads him to make consistently accurate decisions. Rather than just "marking time" and "filling squares", individuals who are professionally vital serve as the mainspring for the institution.

Using Erikson's (1968:91-141) developmental sequence we would predict professional vitality would be the result of successful resolution of successive crises in favor of trust, autonomy, initiative, industry and identity. Trust is the "fundamental prerequisite of vitality." An organization may engender a sense of trust in an individual by that kind of administration which in its quality combines sensitive care of individual needs and a firm sense of personal trustworthiness within the framework of the institution. Also, the institution must present itself to the individual in a reasonably coherent and consistent manner. Autonomy is a sense of free will which is tempered with a recognition of some essential ways in which the individual must conform. The extent to which the individual perceives autonomy is the extent to which the individual accepts responsibility for who he is and where he is going both within the organization and in his life. Next, the individual must develop a sense of initiative which can be likened to the surplus of energy which allows the individual to forget failures quickly and approach new areas that seem desirable. Initiative emphasizes future possibilities and is the essence of personal and professional growth. Initiative leads to the development of what Erikson calls a sense of industry, that is, viewing work as meaningful, important and significant and developing a sense of competence. Trust, autonomy, initiative and industry all are essential to the development of identity, the final important step in the development of professional vitality. Identity has two major characteristics: an internal consistency of values, interests and abilities; and a strong sense of continuity between what one has done in the past with what one plans to do in the future. Thus, to the extent that individuals are able to resolve successive crises in favor of these qualities of trust, autonomy, initiative, industry and identity, they will have the characteristic described above as professional vitality.

Persons with professional vitality are important to an organization because of the productivity and innovation they provide, but perhaps their greatest contribution to the well being of the institution is in what they have to offer their contemporaries and subordinates through mentorship. By returning integrity, generativity, and intimacy to the organizational

environment, professionally vital persons engender trust, autonomy, initiative, industry and identity in those who follow.

In contrast to the individuals with a sense of professional vitality, other individuals have been unable to successfully resolve these successive crises. Unable to experience growth in their environment, they seek relief by withdrawing from the interactional arena. The two major types of withdrawal are extra-organizational and intra-organizational. Extra-organizational withdrawal is commonly called "attrition" because the individual physically removes himself from the institution. Typically, the person choosing this alternative simply continues his quest for personal growth and vitality in a different institution. This type of withdrawal has been much publicized and widely studied. The other type of withdrawal, although equally damaging to the institution, has been largely ignored or overlooked. Intra-organizational withdrawal resembles what Karl Marx labeled "craft-idiocy." Choosing this alternative, individuals often hide their alienation and despair behind a show of disgust and a chronic contemptuous displeasure with the institution in which they, for many reasons, still physically remain (Erikson, 1968:140). In contrast to the person who departs the institution in search of growth opportunities, the craft-idiot accepts his fate and gives up any hope for growth and vitality. Choosing between these two types of withdrawal behavior is often simply a matter of individual needs and perceptions of available alternatives rather than the results of differential crises resolution. Because both of these are detrimental to the institution, an understanding of their developmental process is also important.

These two types of withdrawal, although viewed as being very different, have common developmental factors: mistrust, doubt, guilt, inferiority, and identity confusion. Just as trust forms the basis for the development of professional vitality, mistrust sets the stage for withdrawal. A perception of having been deprived, abandoned, and divided (losses of pay and benefits, faith in leadership, and institutional identification) can all contribute to an underlying sense of mistrust. The failure of the institution to back the individual in his wish to stand on his own feet results in a sense of doubt both in self and the institution, especially if a person has been prepared in childhood to expect personal autonomy, pride, and opportunity. Having developed a sense of mistrust and doubt, the individual next avoids initiative in favor of a sense of guilt over not having progressed toward his professional and personal potential. Burdened with feelings of mistrust, doubt and guilt, it is probable that the individual will resolve the next crisis with his environment toward a feeling of inferiority. According to Erikson, this outcome is enhanced in an organization which has extensive specialization or which emphasizes what the individual cannot do rather than what he can. If mistrust, doubt, guilt and inferiority have been the results of previous crises resolution, the final interactional crisis is likely to lead to identity confusion rather than a sense of identity. Collectively these five resolutions lead to institutional withdrawal.

Throughout this developmental process, individuals have been drawn toward either extra-organizational or intra-organizational withdrawal. The individual who will eventually leave the organization starts exploring alternatives early. Each crisis in turn is approached as a decision

between staying in or getting out rather than in terms of growth and vitality or atrophy and withdrawal. To this extent his eventual exit from the institution becomes a self-fulfilling prophecy, and he may continue his search for vitality in a different environment. This process may also stem from an initial mismatch between the individual and his specific environment. Not every individual can develop professional vitality within every institutional environment.

In contrast, intra-organizational withdrawal often begins with a very good match between the individual and the environment and is typified by a high degree of individual "commitment." The needs of the institution totally dominate the individual's needs. This intense "commitment" is characterized by individuals who do whatever they are told and go wherever they are sent. They fill all the required squares. Although some of these individuals do enjoy some initial institutional success, their neglect of personal and professional growth results in eventual withdrawal because career development without personal growth invariably results in a lack of congruence and identity confusion.

The two types of withdrawal also result in different influences on the interactional arena. Those who have withdrawn from the institution return an aura of isolation to both their contemporaries as well as those who follow. By providing an example of an alternative opportunity for growth, they may also serve to draw others away from the institution. The damage done by those who withdraw within the organization is less obvious but perhaps more serious. They imbue the interactional arena with a sense of stagnation and despair. Their despair is often hidden behind a show of disgust or a chronic contemptuous displeasure with the institution. They show an obsessive need for pseudo-intimacy provided by hollow traditions and flood the interactional arena with stagnation, boredom and interpersonal impoverishment.

In summary, interaction between the individual and the environment leads to successive crises which may be resolved either in favor of professional vitality or organization withdrawal. The value of this model is it places primary emphasis on developmental growth rather than mere physical presence. The current emphasis on such things as commitment and attrition, characterized by such sentiments as "all the way in or all the way out," emphasize only the horizontal aspects of this model and serve only to cloud the issue of the development of professional vitality. Although numbers are important in discussing attrition, the major emphasis should be on quality rather than quantity. By implementing those institutional changes which engender trust, autonomy, initiative, industry and identity and better preparing individuals to successfully resolve these successive crises, the organization will directly enrich the interactional arena which will lead to professional vitality. Both attrition and intra-organizational withdrawal will be reduced as a by-product of these measures.

APPLICABILITY

As an initial check of this model, survey and interview data collected by Frank R. Wood (1980) from 40 junior officers throughout the Air Force, was examined. Fifteen Likert scaled questions were selected which reflect a sense of trust, autonomy, initiative, industry or identity. Answers

reflecting the highest sense of vitality were given a value of 5 while those which reflected the greatest withdrawal were scored 1. These scores were then summed to show a composite vitality index (CVI). The distribution of CVI's ranged from 21 to 68 with a mean of 48.4, a median of 51.00 and a standard deviation of 10.94. An examination of the relationships between each of these variables and professional vitality showed high correlations ranging from .54 to .84 with significance $<.001$. Those cases with CVI's that were greater than one standard deviation above the mean ($CVI \geq 60$) were then labeled VITAL (V) and those cases with CVI's more than one standard deviation below the mean ($CVI \leq 37$) were labeled WITHDRAWAL (W). Finally, the individual groups were examined for similarities and differences across a wide range of variables contained in the data base. Selected similarities and significant differences will be discussed.

Similarities between the groups reflect common characteristics which therefore cannot be used to explain differences in professional vitality. Both groups felt they had a relatively comfortable life style, had experienced PCS moves approximately every two years, felt their spouses supported their career and were evenly divided on the question of whether or not pay and benefits were adequate. Almost unanimously both groups saw themselves as professionals and almost all had received their commissions from OTS or ROTC. In response to questions regarding stated career intent, the modal response from both groups was "20 years and then wait and see." Although the withdrawal group did show slightly less commitment overall, they reflected unanimous institutional orientations as opposed to the vital group in which 29% indicated an occupational orientation.

Differences between the two groups were significant in two primary areas. The majority of the vital group felt they had had "standard" careers (71%), had not experienced a major change in career goals (71%) and had received a regular commission (71%). In contrast, the majority of the withdrawal group felt they had not had a standard career (71%), had experienced a major change in career goals (71%), and had not received a regular commission (71%). Officers in the vital group were generally in support specialties whereas the withdrawal group was a mixture of flying and support officers. On a 5 point scale, average career satisfaction for the two groups was very close (V=3.6, W=3.2), but there was a great difference in present job satisfaction (V=4.1, W=2.3).

The most significant differences in these groups had to do with relationships with supervisors and the development of a mentor relationship. In the vital group, 84% characterized their relationships with their immediate supervisors as "close and friendly." All of the vital group had experienced "mentoring" during their career and felt it had been very important to their progression. Although most of the withdrawal group also felt "mentoring" was important, only two of the seven had received it and then it was in a training situation. In general, the withdrawal group was not as close to their supervisors. Most (71%) characterized their relationships "friendly but not familiar."

Although we recognize the limitations of using a non-random sample, we feel the results shown in this small widely varied sample indicate a need for further study of this model and its implications. It appears there is a high correlation between professional vitality and a sense of trust, autonomy, initiative, industry and identity. Our analysis shows the impact

of the supervisor, in developing and maintaining a mentoring relationship with junior officers, clearly is critical to the development of professional vitality. Our analysis also provides support for the idea that some individuals fail to become vital and withdraw within the organization (craft-idiocy). We cannot over-emphasize the seriousness of the threat that this outcome poses to the well being of the institution. Whereas the institution is enriched by professionally vital individuals whether they stay for 5, 10, 15 or 20 years, the longer a "craft idiot" remains the more harm he does to the interactional arena for those who follow.

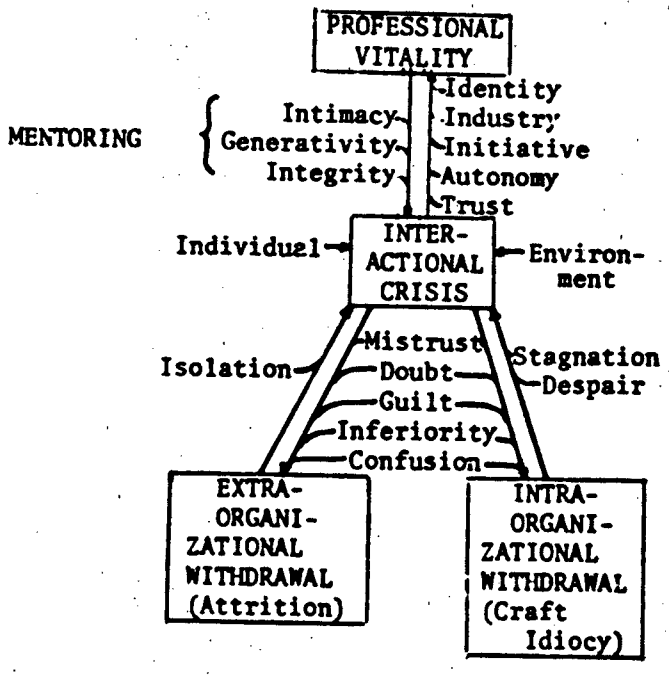


DIAGRAM 1

REFERENCES

Erikson, E. H. Identity, Youth and Crisis. N. Y.: Norton, 1968.

Hall, D. T. Careers in Organizations. Pacific Palisades: Goodyear, 1976.

Super, D. E., Starislevsky, R., Matlin, N. and Jordan, J. P. Career Development: Self-Concept Theory. N. Y.: College Entrance Exam BD, 1963.

Wood, F. R. "Air Force Junior Officers: Changing Prestige and Civilianization." To be published in Armed Forces and Society, Spring 1980.

Cadet Opinions of the Best and Worst Aspects of the USAFA Academic Program[1]

William H. Clover & David B. Porter
Dept. of Behavioral Sciences & Leadership
USAF Academy, Co 80840

Abstract

As part of a survey concerning USAFA cadet attitudes toward the academic program, cadets were asked to describe the two best and two worst aspects of the academic program. 1233 of 1846 cadets provided written responses. Results were content analyzed. Aspects most frequently seen as "best" about the academic program included: faculty, quality of education, class size and diversity of curriculum. Worst aspects included: time pressures, extent of curriculum and competition among cadets. Results are discussed in terms of differences between classes and one special group of cadets.

A fair amount of data has been collected concerning the "climate" at various civilian institutions (Walsh, 1973; Pace & Stern, 1958). However, very little has been done to characterize the academic climate at military academies. This paper describes one aspect of a major survey conducted at the Air Force Academy in November 1978. The overall purpose of this survey is to begin conceptualizing the academic "climate" at the Air Force Academy.

METHOD

ADMINISTRATION & SAMPLE. A 69 item questionnaire[2] was administered in November 1978 during class time to a large cross-section of the cadet wing. Table 1 depicts characteristics of the sample. A total of 1846 cadets responded to the structured part of the survey (approximately 40% of the cadet wing) while 1233 cadets provided written responses to the open-ended questions dealing with the two best and two worst aspects of the Academic program. Table 1 indicates that the respondents are a representative cross-section of the cadet wing.

All students took the survey during classtime with the exception of the cadets labeled "Grassroots." These students were not at the Academy during the administration of the survey. Grassroots participants completed the survey after Thanksgiving break approximately one week after the original administration of the survey. Grassroots cadets represent a

[1] Opinions expressed in this paper are solely those of the authors and do not necessarily represent those of the US Air Force Academy nor the US Air Force.

[2] Copies of the questionnaire may be obtained from the authors.

group of competitively selected cadets from the upper three classes who were helping to increase awareness of the Academy and its program in their home states over the Thanksgiving vacation. Criteria for their selection are shown in footnote c of Table 2. Grassroots cadets were included in the questionnaire sample after the original administration because it was believed that similarities and differences in the responses of this select group compared to the rest of the cadet wing would be instructive.

Table 1

Number and Percentage of Written Responses Compared to Total Survey Response by Group

Group	Number Corresponding to Survey	Number Providing Written Responses	%[a]
Freshmen	708	351	49.6
Sophomore	554	345	61.7
Junior	379	197	52.0
Senior	307	167	54.4
Grassroots[b]	201	146	72.6
TOTALS	1345	1233	56.8

[a] reflects the number of written responses relative to the total number of responses by group.

[b] "Grassroots" refers to a group of Sophomore, Junior and Senior cadets who were selected based on grade point average and military performance to increase community awareness of the Air Force Academy in their hometown communities.

CONTENT ANALYSIS OF WRITTEN RESPONSES. Written responses dealing with the best and worst aspects of the academic program were content analyzed using the following procedures: After the questionnaires were all turned in, the first author and four other officers each independently read a separate sample of completed questionnaires and developed a set of content categories into which the responses of the cadets seemed to fall. Second, all five officers agreed on a common set of categories as well as examples of statements which fit those categories. Third, the written responses were divided among the five officers to determine a frequency count within each category by cadet group. All statements which could not be categorized were held out. Fourth, the five officers met again to resolve the categorization problems which individual officers had problems with. Once these issues were resolved the first author completed a frequency count of all the responses by category by class.

There is one major weakness in this content analysis procedure. There was no check for inter-rater reliability, i.e., would different raters have put the same comment in the same categories? Operational time pressures precluded taking additional time with the content analysis than was taken. Hopefully, the procedures whereby any ambiguous statement was held out for a group decision limited the problems.

RESULTS & DISCUSSION

Table 2 highlights the most frequent comments by topic and by cadet grouping. For purposes of brevity, only the five most frequently mentioned topics in the "best" and "worst" categories are shown and discussed.

Table 2

Percentage of Written Responses Reflecting Best and Worst Aspects

By Group and Topic[a]

Topic	Group[b] 82	81	80	79	GR[c]
	Best Aspects Most Frequently Mentioned				
1. EI from Faculty	39%	37%	2-%	20%	38%
2. Variety/Diversity/Broad Background of Courses in General	27%	21%	32%	33%	45%
3. Class Size/Student Teacher Ratio	18%	13%	10%	12%	17%
4. Quality/Preparation/ Knowledge of Faculty	17%	13%	29%	23%	32%
5. Quality of Education	14%	17%	17%	18%	13%
	Worst Aspects Most Frequently Mentioned				
6. Time Pressures	35%	37%	31%	25%	40%
7. Too Many Core Courses	14%	24%	25%	14%	21%
8. Emphasis on Grade/Gamesmanship/ rather than Knowledge/Learning	9%	20%	19%	22%	29%
9. Acad Load Too Great/Too Many Courses In General	10%	10%	12%	8%	14%
10. Competition Among Cadets	8%	9%	8%	9%	10%

a. Percentages reflect frequency of occurrence of response and therefore do not add to 100%.

b. Class reflects anticipated year of graduation (i.e., 79=seniors, 80=juniors, etc.). n for each group is reflected by Written Response Column of Table 1.

c. Grassroots (GR) were treated as a separate group. Competitively selected by state to increase community awareness of Academy and programs. Criteria: '79 & '80: 2.75 CWA equivalent to a 2.5; '81=3.0 equivalent to a 2.75. Military standing was also considered.

BEST ASPECTS. Examples of cadet comments dealing with what they perceive as the best aspects of the academic program are: "The amount of extra instruction available to cadets. I feel that the instructors provide help and are willing to help with any problem we have...it still amazes me that faculty members care enough to give out their home phone numbers." "...the chance to take a wide variety of courses that most people at a civilian college either wouldn't take, or aren't offered." "the broad type background is a terrific mix of science and humanities." "Small classes, more of an interpersonal relationship between instructor and student compared to other schools." "The quality of the education we receive, and the high calibre of the instructors."

Three of the five categories of "Best Aspects" in Table 2 deal with the faculty in some fashion as a "best" aspect of the academic program. The most frequently mentioned item was "Extra Instruction" (EI) from the faculty. Almost 40% of the freshmen and sophomores felt that EI was one of the best aspects. The proportion of people mentioning this from the junior and senior classes, although still high, drops off considerably

(24% and 20%) while the percentage of Grassroots cadets stays relatively high (38%). It seems that this response is a function of need. Being new to the academic program, freshmen and sophomores may take advantage of EI more frequently because it is a genuine assistance toward maintaining their grades. Juniors and seniors may have adapted to the academic program and therefore feel less need for EI. Additionally, juniors and seniors have chosen academic majors in their area of interest and this may offset their need for extra instruction. Additionally, many poorer academic students have left the Academy and the remaining cadets can successfully navigate the academic program without assistance. It seems to be anomalous that the Grassroots students would mention EI so frequently. Perhaps this is because many of them find EI helps them maintain their grades. Another possibility is that EI also represents a way to get to know faculty members and find out about the Air Force. In that event, what starts as EI may become counseling or simply friendship.

A second area related to the faculty is shown on line 4 of Table 2. These comments dealt with the quality, class preparation and knowledge of the faculty. There is a generally increasing trend from the freshman (Class 82) through the senior year (Class 79). Perhaps the greatest factor causing this is the fact that cadets generally begin work in their academic major in their junior year. This experience allows them to interact with the faculty in their area of interest on a more frequent basis. Overall, Grassroots students mentioned the quality of the faculty more frequently than any of the classes individually. Perhaps this is because they have interacted more with the faculty (recall their reaction to EI) or because they have "bought into" the system more than other cadets.

The third area which is related both to the faculty and the institution deals with the "quality of education at the Academy" (Line 5, Table 2). In general, the proportion of people who mentioned this aspect of the academic program remained fairly constant (i.e., 16%-18%) across the four classes. The confusing percentage comes from the Grassroots group. Only 13% mentioned this category. On reflection this could be the result of any one of several reasons. First, the unreliability in the scoring system. Second, it's possible that this area seemed redundant to other comments and was therefore not used by Grassroots students. Third, and most controversial, is the possibility that these students who rank high both militarily and academically actually don't feel as positively about the "depth" of their education. This area deserves further study.

In contrast, Line 2 of Table 2 shows a very positive set of responses which dealt with the variety, diversity and broad background of courses in general. Here we found a relatively strong response across all classes. While there was a slight upward trend from freshman through senior years, this area was the second most frequently mentioned comment dealing with best aspects of the academic program. Note that 45% of the Grassroots cadets made comments which fell in this area. This may offset comments made earlier about the possibility that they may perceive a lack of depth. It is reasonable to assume that these students value the broad range of courses they take while not being pleased that they don't go into greater depth.

The final category which fell under best aspects of the academic program reflects responses which dealt with class size and student teacher ratio. Here we see a drop from the freshman year and then a level frequency from the sophomore through senior year. It is possible the drop is due to adaptation. While it is a unique aspect of the Academy that classes normally do not exceed 20 students, the novelty of small classes becomes an expected feature after the first year. Here again, though, we see that Grassroots cadets responded more frequently (16%) to this feature of the academic environment than their classmates.

Overall, cadets were able to highlight several areas they thought were "best" about the academic environment. These centered on the faculty, the curriculum and the size of classes. As some might expect, Grassroots students seemed generally more positive about these same aspects of the academic program than other cadets but in one area were not as enthusiastic. We now turn to what cadets stated are the worst aspects of the academic program.

WORST ASPECTS. Examples of cadet comments concerning "worst aspects" of the academic program are: "We need more time. The academic program is like taking a drink of water from a fire hydrant." "Unfortunately, I feel I'm studying more for grades than knowledge...there's just enough time to "spec" the material, but not enough time to think about it really seriously." "Competitive 'cutthroat' feeling it instills; i.e., 'hope he does bad so mean is low'." "I get the impression that...the academic side of the house feels cadets will automatically become well rounded and intelligent by having 180-187½ credits crammed down their throats...". "The preponderance of core courses, none of which are offered as options to the cadet...".

Table 2 shows that the item most frequently mentioned as a "worst aspect" deals with "time pressures." The proportion of people mentioning this within the classes ranged from 35% of freshmen to 25% of the seniors. While there is a decrease over time, it still remains the most frequently mentioned category. It is interesting to note that Grassroots cadets (Sophomores, Juniors and Seniors) mentioned time pressures more than anyone else (40%). This may be because people in the Grassroots category literally do have less time than any other group. It is possible that Grassroots people participate in more Academy activities and therefore have more of their time committed.

Lines 7 and 9 of Table 2 highlight two areas closely related to the idea of time pressures. These deal with the ideas that: (a) there are too many core courses and that the academic load is too great with too many courses in general. Looking at Line 2 (too many core courses), we see that sophomores (Class 81) and juniors (Class 80) mentioned this most frequently (24% and 25% respectively). Freshmen (Class 82) and seniors (Class 79) mentioned it 14% and 16% respectively. One rationale for this chain of events might be that freshmen are so busy they don't know what to expect; therefore, the fact that they are taking only core courses doesn't dismay or surprise them. On the other hand, seniors are taking mainly major courses and therefore don't feel the pressure of so many core courses. Of course, by this time seniors may also be feeling that "it wasn't so bad."

After all, they have been successful! Grassroots cadets mentioned this issue 21% of the time and could be reflecting on the same issues as their sophomore and junior colleagues or they could be stating their own desires to take more electives or major courses. Whatever the reason, "too many core courses" was an important issue in the cadet comments.

Line 9 of Table 2 reflected the other area which seemed to be related to time pressures. These comments were summarized under the category "academic load too great with too many courses in general." Here we see a generally consistent picture over the four classes with a slight upswing in the junior year (12%) and a dip in the senior year (8%). Grassroots cadets mentioned this item more than any of the other classes taken separately (14). In summary, three of the five areas mentioned as worst aspects of the academic program dealt with the constraints on a cadet's time. The combined totals for these three areas remain a relatively constant 60% of responses by class and 80% of the Grassroots cadets. The legitimacy of this concern perhaps needs an independent source of observation and evaluation. That source could be the observations of the distinguished visiting professors who spend a year at the Academy or another source who could make comparisons between all the military academies, while understanding their history, purpose and objectives.

Another area which drew several comments as one of the worst aspects of the academic program dealt with what cadets perceived as an "emphasis on grade/testmanship rather than knowledge/learning." Line 8 of Table 2 indicates a significant increase in this concern from the freshman to the sophomore year. From the sophomore through the senior year was relatively constant although there was a slight increase among the seniors (23%). Grassroots cadets mentioned this issue more than cadets in any single class (25%).

The final issue shown in Table 2 deals with "competition among cadets." A consistent 8–10% of each class and 10% of the Grassroots respondents listed this issue as one of the worst aspects of the academic program. In general, cadet comments indicated that competition was seen as a divisive influence on their relationship with other cadets.

CONCLUDING COMMENTS

The data from these written comments indicated several areas which might possibly be explored further in a more systematic manner. The faculty stood out as the major strong point of the academic program as well as various aspects of the overall curriculum. On the other hand, issues dealing with time pressures, size of the core curriculum, total number of courses and competition among cadets stood out as the most negative aspects of the academic program. Future research should attempt to explore the relative importance of these issues in terms of their relationship to attrition and overall satisfaction. It is not suggested that cadets be the sole criterion for what should or should not be included in their academic program. Nor is it suggested that cadets not be listened to if they can't provide the cure for their own ills. It is suggested that perhaps the symptoms they describe may be indicative of a problem to be treated. Only further study and careful concern will determine that.

Table 1

Number and Percentage of Written Responses Compared to Total Survey Response by Group

Group	Number Responding to Survey	Number Providing Written Responses	%[a]
Freshmen	708	381	53.8
Sophomore	554	342	61.7
Junior	379	197	52.0
Senior	307	167	54.4
Grassroots[b]	201	146	72.6
TOTALS	1846	1233	66.8

[a]% reflects the number of written responses relative to the total number of responses by group.

[b]"Grassroots" refers to a group of Sophomore, Junior and Senior Cadets who were selected based on grade point average and military performance to increase community awareness of the Air Force Academy in their hometown communities.

Table 2

Percentage of Written Responses Reflecting Best and Worst Aspects By Group and Topic[a]

Topic	Group[b] 82	81	80	79	GR[c]
	Best Aspects Most Frequently Mentioned				
1. EI from Faculty	39%	37%	26%	20%	38%
2. Variety/Diversity/Broad Background of Courses in General	27%	31%	32%	33%	45%
3. Class Size/Student Teacher Ratio	18%	13%	12%	12%	16%
4. Quality/Preparation/ Knowledge of Faculty	17%	13%	29%	23%	32%
5. Quality of Education	16%	17%	17%	18%	13%
	Worst Aspects Most Frequently Mentioned				
6. Time Pressures	35%	28%	31%	25%	40%
7. Too Many Core Courses	14%	24%	25%	16%	21%
8. Emphasis on Grade/Testmanship/ rather than Knowledge/Learning	9%	20%	19%	23%	25%
9. Acad Load Too Great/Too Many Courses In General	10%	10%	12%	8%	14%
10. Competition Among Cadets	8%	9%	8%	9%	10%

a. Percentages reflect frequency of occurrence of response and therefore do not add to 100%.

b. Class reflects anticipated year of graduation (i.e., 79=seniors, 80=juniors, etc). n for each group is reflected by Written Response Column of Table 1.

c. Grassroots (GR) were treated as a separate group. Competitively selected by state to increase community awareness of Academy and programs. Criteria: '79 & '80; 2.75 GPA waiverable to a 2.5; '81=3.0 waiverable to a 2.75. Military standing was also considered.

AGES: From Development to Implementation

Earl S. Stein
US Army Research Institute for the Behavioral and Social Sciences

Abstract

The training of air defense personnel has lacked realism and participant motivation. Air-Ground Engagement Simulation (AGES) has been developed to provide realistic combat simulation coupled with an implementation program. AGES was field tested in USAREUR by comparing AGES trained squads (n=6) against similar squads with traditional field training (n=6). Performance was assessed by using observational behavioral ratings. Results indicated a training advantage for AGES squads. The implementation program has stressed media dissemination of AGES concepts and the development of a Mobile Training Team (MTT). Soldier acceptance of the MTT approach was evaluated during field exercises. Analyses of responses from participant air defenders and air crews indicated that soldiers felt AGES was realistic and beneficial training.

Collective training in the US Army has traditionally followed a pattern of highly structured preplanned exercises which are based on combat scenarios. Combat soldiers are taught through experience that everything happens according to plan, and what they do as individuals or their unit does as a team has very little influence on the outcome of a tactical exercise. Because results are not systematically related to behavior, soldiers find it very difficult to acquire tactically proficient skills and coordination. The simulation of the rewards and costs of combat are seldom available in traditional collective exercises.

The techniques of Engagement Simulation (ES) have evolved over the past eight years to provide a positive alternative for unit training. ES emphasizes realism through the simulation of weapons' firing signatures and through the assessment of casualties according to specific, believable rules. When a soldier is informed that he has been "killed" in an ES exercise, he knows that he was in someone's sights and that he "died" because of what he did and/or because of what the enemy did.

ES has developed from the training of rifle squads up to and including combined arms company teams, (TC 71-5 REALTRAIN, US Armor School, 1975). The training of Air Defense personnel was a relative latecomer to ES. Air-Ground Engagement Simulation (AGES) was an outgrowth of REALTRAIN and emphasized the same principles.

The goals of this research effort were two: First, demonstrate the training effectiveness of AGES over traditional methods; and second, involve the training research community in the implementation of their product. The sharing of the responsibility for implementation by research scientists is a novel step beyond their traditional role.

Method-Phase 1 Validation

A validation effort was accomplished in the US Army Europe during the summer of 1978. The procedure which was discussed in detail by Erwin & Stein (1979) will be briefly summarized here.

Participants included members of four Chaparral Squads, four Vulcan Squads and four Redeye Teams. Half of this group was assigned randomly to an AGES training condition and the other half was provided "conventional" training. Repeated measures for eight exercises were run for each set of squads. In the AGES group, ES exercises were run with the air defense forces defending critical assets against opposing force scout and attack helicopters. Casualties were assessed using a probability system which was linked to the lethality of the individual weapons system (both air defense, and attack helicopter), which was fired in any given engagement. The conventional training group was given the same missions and was opposed by the same aircraft. However, neither casualty assessment nor signature simulation were accomplished. Another distinctive feature of AGES was the after action review which allowed a free interchange of information by participants on both opposing sides. This was provided to the AGES but not to the conventional group.

Squad or team performance of ADA personnel in both treatment groups was evaluated using a behavioral checklist/rating form which was developed out of the technical literature and through the use of subject matter experts. A composite score was developed for each force on a daily basis.

Results-Phase 1

The effectiveness of AGES training as compared to the traditional approach appeared to be system specific. Crews on both Chaparral and Vulcan AD weapons performed better with AGES training. There appeared to be no difference for Redeye Teams, however. When data were pooled across the three systems and crew performance change was calculated as the % of performance improvement from the first day of training to subsequent days, the advantages of AGES were apparent (Table 1).

Table 1

Crew Performance Change (%) From the First Training Day

Training Type	Training Day		
	2	3	4
AGES	38%	46%	11%
Conventional	-20%	-11%	-07%

While AGES was superior every day, its lead diminished especially on the fourth day. It was apparent that the conventionally trained squads "lost ground" after the first day of training and only gradually recouped. It would have taken at least another day of training for them to reach their level of performance on the first day.

Method-Phase 2 Implementation

Once a training system is developed and some training effectiveness data have been collected, the research community usually documents its' findings and goes on to another challenge. However, the developers of AGES decided to take their responsibility one step further. This required establishing a strategy to encourage use of the AGES training system.

The first step was to convince the air defense community that this training was important enough to establish a mobile training team to serve as a credible nucleus for AGES expertise. Once personnel were assigned to the mobile training team, the next step was to provide them with draft training materials to include programs of instruction and the detailed outline of a training circular. Draft rather than finalized materials were provided, because we believed that implementation would be easiest if the training team participated fully in developing the program. This would then be viewed as an internally developed product of the air defense rather than the research community. Also, the mobile training team would be more likely to take pride in a training product which they viewed as their own.

Coupled with the development of the training team and its materials was an attempt to market AGES in the appropriate media channels. An article was published on the program in the Air Defense Magazine which is a major communication channel in the Air Defense community (Erwin & Stein, 1979). The mobile training team took it upon itself to send a letter of AGES introduction to every Chaparral/Vulcan battalion in the US Army.

None of the members of the mobile training team had ever seen or participated in an AGES exercise. The next step in this phase was to run implementation trials to evaluate the effectiveness of the team and the materials which they developed. In August of 1979, a series of exercises was run to accomplish this goal.

A composite air defense platoon was organized, consisting of two Chaparrals, two Vulcans, and two Redeyes. This platoon was opposed by elements of an air calvary troop consisting of two Scouts and two Attack helicopters. The purpose of these exercises was not revalidation but rather to train and assess the credibility of the mobile training team as viewed by training participants. Given these objectives, data collection consisted of attitude/opinion surveys of participants after a series of five days of AGES exercises.

Results-Phase 2

Table 2 presents a summary of the questionnaire responses by air defense personnel. Average responses to the questions and response frequencies appear very favorable to AGES training. The majority of personnel felt better trained after AGES exercises and were very enthusiastic about their experiences. All responses were on an 8 point scale, where 8 indicated the strongest level of agreement. Response frequencies are reported in Table 2 and in subsequent tables. The left column (≤ 4) represents the number of people who disagreed with the statement to some degree and assigned a scale value of 4 or less. The right column (≥ 5) represents the frequencies of agreement responses, 5 or higher. One survey item was stated in such a way as to generate a potentially low response. This was the statement "AGES was not realistic training." A mean response of 2.74 to this item indicated substantial disagreement with the statement and also supported the assumption that respondents were reading and answering each item individually.

When air defense leaders were asked about AGES' effect on their leadership, they indicated that AGES encouraged a leader to keep his subordinates trained (mean 7.27) and gave him the opportunity to see what his people could do in simulated combat (mean 7.64).

Table 2
Summary of Questionnaire Response for ADA participants (N=18)

Questions	Mean	SD	Response Frequency ≤4	Response Frequency ≥5
AGES helped my squad or team learn to work together as a team	6.56	2.20	4	16
AGES exercises were realistic and gave me a better feel for the sights and sounds of battle	6.11	2.27	5	13
I feel better trained to perform my combat duties	5.89	2.59	5	13
I enjoyed AGES training	6.26	2.38	4	14
AGES increased my confidence in my weapons and their effects	6.00	2.03	4	14
AGES was not realistic training	2.74	2.18	14	4
AGES made me more aware of the physical dangers of combat	6.32	2.60	4	14
AGES was worthwhile to all the members of my squad or team	5.89	2.73	5	13

Response Scale

Strongly Disagree	1	2	3	4	5	6	7	8	Strongly Agree

Table 3
Summary of Questionnaire Response for Control Personnel

Question	Mean	SD	Response Frequency ≤4	Response Frequency ≥5
The overview of AGES presented by MTT personnel gave me a clear understanding of AGES	6.71	1.50	1	6
After training by the MTT before the AGES exercises, I understood how casualty assessment works	5.57	2.70	2	5
The training by the MTT prepared me to install and reload the Signature Simulators	4.15	3.15	3	3
The MTT's explanation of how to fill out the control forms was clear and understandable	7.29	1.89	1	6
As part of the control system I still learned something about Air Defense techniques	8.0	0	0	5

Response Scale

Strongly Disagree	1	2	3	4	5	6	7	8	Strongly Agree

While the results to this point seem to infer indirectly that the military training team was doing its job well, responses from control personnel who actually ran the exercises reflected directly on the MTT. Table 3 summarized the controllers responses. The MTT provided to them a clear understanding of AGES principles and the use of control system forms. The initial training in the maintenance of the signature simulation hardware was a weak area as indicated by responses to the third question (Table 3). This did not result in any practical problems in the field, however, and controllers with the assistance of crew members were able to keep the simulators operable.

Discussion

The validation effort in Europe provided enough favorable evidence for the potential of AGES to make it a marketable program. Marketing efforts have resulted in calls and letters to the mobile training team from virtually all Army Air Defense units. The original implementation concept called for the funding of MTT visits by the requesting unit. This proved unfeasible in the short run due to programming considerations for temporary duty funds. However, the MTT is attempting to secure centralized funding for future site visits. While waiting for this problem to be resolved they are attempting to put together a "Do it yourself" package which can be mailed to requesting units.

The current state of the art with AGES technology requires casualty assessment to be accomplished manually by the Control System using probability and distance estimation between target and firing weapon. Maneuver units will in time have a laser ES System referred to as MILES (Multiple Integrated Laser Engagement Simulation) which vastly simplifies the casualty determination and reduces the number of people required as controllers. This introduces a question into air defense training. Should trainers go ahead with what they have available now or wait for a laser based system? The answer is very clear. Since a MILES-AGES is scarcely on the drawing board in comparison to that for maneuver units, the time before eventual fielding is measured in years. Air defenders can not afford to wait, and should make use of what is currently available--AGES, an innovative and motivating training system.

References

Erwin, D.E, and Stein, E.S. AGES: Realistic tactical training for Air Defenders. Air Defense Magazine, July-Sept 1979.

Stein, E.S., and Erwin, D.E. Air-Ground Engagement Simulation (AGES): Realistic and effective training for air defense personnel. In Proceedings of the first Inter-Service/Industry Training Equipment Conference (Tech Report IH-316), Naval Training Equipment Center, Orlando, Florida, Nov 1979.

US Armor School. REALTRAIN: Tactical training for combined arms elements (TC 71-5). US Armor School: Ft. Knox, KY, 1975.

Development of a Simulator Certification Methodology for Strategic Air Command

K. Ronald Laughery, Jr.
Calspan Corporation
Buffalo, New York 14225

Major George M. Houtman
93 BMW/DO5, Instructional Systems Development Division
Castle AFB, California 95342

Abstract

The approach used to develop the Simulator Certification (SIMCERT) Program for the Strategic Air Command (SAC) is presented. The methodology developed and a test plan for validating the methodology are also included.

Background

Air Force Regulation 50-11 (October 1977) requires that all new and existing aircrew training devices (ATD) undergo a process called Simulator Certification (SIMCERT). The regulation mandates completion of initial certification during follow-on test and evaluation (FOT&E), and intermittent assessments of training effectiveness throughout the ATDs' life cycle. AFR 50-11 defines SIMCERT as "The process of verifying (according to AFR 60-) the specific aircrew tasks that can be a) effectively trained in an aircrew training device and b) credited toward training reqirements as established in each respective 50- and 51- series publication." (para. 1.d.) In short, SIMCERT seeks to document how ATDs facilitate the learning of specific skills and permit valid performance assessment of those skills. Each using command must publish a supplement to AFR 50-11 setting up a SIMCERT process that meets its specific needs. This paper describes the methods used to develop an optimum SAC SIMCERT methodology and outlines how to conduct the preliminary evaluation of a candidate approach.

Method

General.

In May 1979, Calspan Corporation and 93 BMW/DO5 at Castle AFB began a two-phase program to develop the SAC SIMCERT process. Phase I, which is finished, established the groundwork for the beginning of a SIMCERT methodology. The three steps during Phase I were 1) literature review, 2) determination of objectives and resource constraints, and 3) development of candidate methodologies. Phase II is presently under way and involves validating the methodologies to ensure that they are feasible and meet the objectives for which they were designed. The following subsections describe the approach and rationale used during Phase I to develop the SIMCERT plan, and the candidate methodologies to be validated during Phase II. A

PRECEDING PAGE BLANK-NOT FILMED

progress report on Phase II is contained in the Discussion section.

Literature Review

The literature review established the theoretical basis for SIMCERT. Attempts to address similar issues were examined to take advantage of successful approaches and avoid documented mistakes.

Literature on learning principles pertinent to simulator effectiveness and evaluation was examined on the assumption that a simulator's utility is affected by its application within an instructional system. Previously, attempts have been made to evaluate the simulator without regard to the instructional system, but this approach is no longer considered acceptable by many training professionals.

Literature on experimental design was examined to determine which experimental methodologies are applicable to certification. The quality of SIMCERT results will depend on how well the transfer of training experiments are designed and controlled. Emphasis was placed on statistical and experimental design literature pertaining to validation and transfer of training experiments. A distinction is made between a simulator's utility as a performance assessment tool (e.g., checkrides), and its utility as a training device (transfer of training). The performance assessment function requires that the simulator be a valid predictor. The individual's performance score in the simulator must predict how well he will do in the aircraft. If simulator scores predict aircraft scores over a wide range of scores, then the simulator is a valid measurement device. If the simulator is to be used only as a training device, validity is not essential. Instead, it must be demonstrated that the simulator produces transfer of training. Transfer of training implies that the simulator will permit a student who is unskilled in the aircraft tasks to obtain these skills more quickly with simulator practice than without it. To summarize, validity refers to the device's ablity to predict performance and transfer of training refers to the device's utility in facilitating learning. It was deemed critical to clarify this distinction in order to arrive at an experimental design which would evaluate the simulators effectiveness in satisfying each of the two requirements.

Finally, recent transfer of training studies were reviewed. It was anticipated that the certification methodology developed for SAC would involve some combination of the techniques used in these studies.

Determination of SAC SIMCERT Objectives and Constraints

While AFR 50-11 generally defines SIMCERT objectives, there is much room for interpretation. Therefore, it was considered essential to clearly define concrete objectives so that the SIMCERT methodology could be designed to specifically address these objectives. These decisions were managerial rather than technical. Consequently, the approach used to determine the appropriate SAC SIMCERT objectives was to delineate all reasonable alternatives and then present them to the appropriate management decision makers within SAC for review.

Essentially, it was determined that any objective addressing issues related to optimizing the training device or training system were

not to be a part of SAC SIMCERT. Instead, SIMCERT would be the examination of the training device, or set of devices, as currently fixed within a given instructional system. Rather than a developmental process, SAC SIMCERT is the process of ensuring that a given training device facilitates adequate training and student performance evaluation within a given training system. The operational test and evaluation (OT&E) process will address the developmental aspects of the simulator within the training system. A further clarification of SAC SIMCERT objectives is included in Laughery and Ditzian (1980a).

Given SAC SIMCERT objectives, a logical question is, "What resources (personnel and material) are available to do the job?" This is accompanied by another question: "What resources are necessary to do the job?" The answers to these questions are management decisions which must be based on technical considerations. The results of the SIMCERT effort will largely depend on the resources applied.

Various approaches were considered to obtain information of the trade-offs that are necessary. One approach considered was to list all resources which would potentially be required. Next, personnel with knowledge of constraints would determine how much of each resource would be available. This approach has two definite drawbacks. First, it seems unlikely that problems could be identified out of context, and without a more global picture of the total amount of resources required. Secondly, the benefits of one method requiring more resources versus another method requiring fewer resources would not be apparent. To overcome this difficulty, the approach selected was to develop three scenarios on how SIMCERT could be conducted and thereby define the resources which would be required. Each scenario represented an approach to SIMCERT and addressed certification for both initial and continuation training.

These scenarios were first reviewed by HQ SAC/DOTP and 93BMW/DO5 to ensure that terminology, context and resources were appropriate. Copies of the scenarios were distributed to other agencies in the command for review. The primary purpose of this wide review was to obtain feedback on the feasibility of the approaches to SIMCERT. The three scenarios were not intended to represent all possible approaches to SIMCERT. Rather, they were a cross section of approaches in terms of resource requirements. Of the 29 copies sent out, 13 were returned with comments. These comments were reviewed and tabulated by Calspan.

In general, the comments and suggestions indicated that the evaluators understood their task. As expected, most comments reflected the particular concerns of the individual reviewers. Significant information was obtained concerning resource constraints. For further information of specific resource constraints, see Laughery and Ditzian (1980a).

SIMCERT Methodologies

Once the SIMCERT objectives were determined and the constraints were identified, the problem remained to match these needs to the best technical approach that would remain within operational constraints. Two methodologies were developed: one for simulator use in initial training, and one for simulator use in continuation training. Both allow for evaluation of a simulator's performance assessment ability and its training

ability. Except for differences in transfer of training evaluation, the two approaches are identical, and contain six steps. Steps 1-3 involve the first evaluation of validity and training effectiveness.

Step 1. Documentation of simulator fidelity. A key component of SIMCERT is to ensure that an ATD maintains approximately the fidelity it had when it was initially accepted. If the fidelity remains relatively constant, greater confidence can be placed in the simulator's training utility. This step establishes the baseline for fidelity.

Step 2. Determination of the way in which the ATD is used in the training program (training application). As in Step 1 above, if the way in which a simulator is used remains constant, there is greater confidence that the simulator's effectiveness will remain constant. This step establishes the baseline for simulator use procedures.

Step 3. Transfer of training evaluation. Clearly, the bottom line of simulator effectiveness is the extent to which training is facilitated through use of the simulator. Consequently, considerable effort will be expended to collect and evaluate data on the transfer of training. Specific experimental designs have been proposed which include a control group and two experimental groups. Also, specific types of student performance data will be collected.

The documentation of simulator fidelity (Step 1) and training application (Step 2) will be more useful if these parameters can be monitored and reassessed during recurrent evaluations. If fidelity and training application do not change, more confidence can be placed in the assumption that all generations of students will undergo the same transfer of training. This approach is simplistic and does not exhaust all factors affecting simulator training effectiveness. However, fidelity and training application are clearly the strongest influences that can be readily monitored. Therefore, recurrent evaluation of the simulator will primarily involve ensuring that simulator fidelity and application remain relatively constant. It will be assumed that training effectiveness remains constant as long as fidelity and training application remain constant.

The next steps (4-6) involve recurrent evaluation.

Step 4. Maintaining simulator fidelity. This step involves inspection of the simulator by experts knowledgeable in the operation of the actual equipment to ensure that fidelity has not changed appreciably.

Step 5. Maintaining simulator use procedures. This step is to ensure that the simulator is still being used in the portions of the training program as originally planned.

Step 6. Monitoring performance of the trainees who have been trained in the simulator. This step closes the loop on recurrent evaluation. It is essential to periodically monitor student performance, but in a less rigorous manner than in Step 3.

Steps 1 and 2, which require the documentation of fidelity and training application, will be performed prior to the assessment of transfer of training. Once this documentation is available, the transfer of training study will be conducted to verify which tasks can or cannot be trained and/or evaluated in the simulator. The tasks found to be adequately trained and/or evaluated should be presented to the proper

agency along with supporting documentation. At this point, the simulator would be certified for these tasks or events.

Next, steps 4 through 6 would be performed on a routine basis as a part of recurrent evaluation. If changes in fidelity of training application are found, an analysis will be conducted to determine for which tasks the simulator should be decertified. This determination could be made by examining the systems used in task performance (fidelity) or the lessons in which the task has been taught (training application).

When fidelity or training application change, and tasks are decertified, it would become necessary to either repair the simulator or its application, or accept the change permanently. If the alteration is repaired, the simulator could be rechecked and re-certified for the applicable tasks. If not, it would be advisable to conduct the initial certification only for those tasks which are affected.

Finally, the monitoring of graduate performance will provide the extra measure of quality control to ensure that the results obtained during initial certification remain valid. If a substantial deterioration in graduate performance is detected, the device will be decertified for the affected tasks until another transfer study can be conducted.

Discussion

At this time, the SIMCERT methodology discussed herein is being applied and evaluated in the KC-135 Boom Operator Part Task Trainer at Castle AFB CA. As problems are identified, the methodology will be modified to better meet the SAC needs.

This evaluation of the methodology will continue on the B-52 Air Refueling Part Task Trainer and Cockpit Procedures Trainers. By applying the methodology on a cadre of simulators, problems can be eliminated prior to application on the B-52 Weapon System Trainer. Certification of a full mission, full crew simulator will certainly provide the ultimate test of the methodology's effectiveness.

References

Ditzian, J. L. and Laughery, K. R. SAC Simulator Certification - Literature Review (Calspan T.M. 9-14.2/1). Calspan Corporation, Buffalo, N.Y.: Calspan Corporation, August 1979.

Laughery, K. R. and Ditzian, J. L. SAC Simulator Certification - Objectives and Constraints (Calspan T.M. 79-14-2/2). Buffalo, N.Y.: Calspan Corporation, February 1980.

Laughery, K. R. and Ditzian, J. L. SAC Simulator Certification - Candidate Methodologies (Calspan T.M. 79-14.2/5). Buffalo, N.Y.: Calspan Corporation, February 1980.

TRAINVICE Analysis of Generic Electronic Maintenance Trainer Concepts

A.J. Josefowitz and J.W. Kochevar
Honeywell Systems and Research Center
Minneapolis, Minnesota 55413

Abstract

The TRAINVICE method of analyzing training device effectiveness was applied to two conceptual approaches to generic electronics maintenance training. A two dimensional approach to training simulation was compared to a three dimensional approach. The intent of the analysis was to determine the generality of the TRAINVICE method and to assess difficulties in its application. The paper presents a brief description of the TRAINVICE method as originally conceived and discusses modifications necessitated in its application to generic trainer concepts. Assumptions and difficulties in using the method are described. Problems encountered in applying the TRAINVICE method to device concepts (instead of existing devices) and to generic trainers (instead of equipment-specific devices) are elaborated.

Recent DoD directives (5000.1 and 5000.2) have stressed the importance of early trade off analyses among the various subsystems within the Weapon System acquisition process. This means that certain support subfunctions of a weapon system such as training must be taken into account when planning the acquisition of a weapon system. For this to be achieved, it is necessary that training requirements be identified early in the acquisition process. An increasingly important part of training requirements analysis is the identification of requirements for training devices. While the early identification of training device requirements is important, the task is extremely difficult. Ideally, device requirements should be identified during the conceptual phase of system development. Alternative training device concepts should be compared in trade-off analyses, and the impact of alternative hardware designs on training device requirements should be incorporated into system hardware trade-off studies. For this to be accomplished, methods of analyzing training device effectiveness are needed. The present study was intended to examine the utility of one recently-developed method for analyzing training device effectiveness.

The TRAINVICE method (Wheaton, Rose, Fingerman, Leonard, & Boycan, 1976b) was applied to two generic electronics maintenance trainer concepts to develop an understanding of user requirements, constraints in using the procedure, its generality, data requirements, and ease of use. The present use of TRAINVICE differed from its initial application. TRAINVICE was originally applied to the problem of evaluating training effectiveness of existing training devices designed to develop equipment-specific skills (Rose, Wheaton, Leonard, Fingerman, & Boycan, 1976; Wheaton, Rose, Fingerman, Leonard, & Boycan, 1976a; Wheaton et al., 1976b). This paper discusses the application of the procedure to alternative devices in the conceptual stage of development. Furthermore, the concepts examined were generic rather than specific skill trainer concepts.

As originally presented (Wheaton et al., 1976a; 1976b), the TRAINVICE procedure required the analyst to evaluate device effectiveness using six

PRECEDING PAGE BLANK-NOT FILMED

rating scales: Commonality of subtasks performed on the device and on the operational equipment, physical similarity of controls and displays on the device and on the operational equipment, functional similarity of controls and displays on the device and on the operational equipment, learning deficits to be overcome by use of the device, subtask learning difficulty, and appropriateness of training technique. Two or more alternative devices would be evaluated on each of these six scales. Scores on the scales would be combined to yield a single training effectiveness score for each device under consideration. Comparison of these effectiveness scores provides the basis for judging the relative effectiveness of the devices.

The following information is required to perform the analysis: (1) lists of tasks to be performed on each device being compared and on the operational equipment, (2) lists of subtasks to be performed on each device being compared and on the operational equipment, (3) lists of controls and displays involved in the performance of each task, and (4) training objectives for each task.

Method

While every effort was made to maintain the original TRAINVICE procedures in analyzing generic trainer concepts, modifications were necessary. The transition from equipment-specific existing trainers to generic trainer concepts as the target of analysis made it difficult to apply some of the original rating scales.

In part, these difficulties were created by the lack of detailed data associated with equipment in conceptual stages of development. However, this has not been viewed as a particularly difficult problem in applying TRAINVICE (Hirshfeld & Kochevar, 1979; Narva, 1979; Wheaton et al., 1976b). For the present analysis, the device concepts under consideration were taken from Type A, Form 1B specifications for two alternative generic electronics maintenance trainer concepts (Heeringa, Koch, Persons, Daniels, Pine, & Squires, 1979; Daniels, Heeringa, Koch, & Pine, 1979). These concepts represented two-dimensional (2-D) and three-dimensional (3-D) trainer approaches lacking detail in specification of physical and functional aspects of controls and displays as required in a similarity analysis. This problem was resolved by assuming maximum capability for the respective 2-D and 3-D concepts, given the constraints contained in the specifications. In other words, the "best" 2-D concept within specifications was compared to the "best" 3-D concept within specifications.

Another factor which necessitated modification of the procedure was associated with the generic nature of the device concepts under consideration. Since the device concepts were unrelated to any particular operational equipment, no operational(rather than training) tasks, subtasks, controls, or displays could be specified in the analysis. The TRAINVICE procedure is to compare these aspects of operational equipment with each alternative device concept. This was not possible in the present analysis. Instead, the two concepts were directly compared against each other.

This particular analysis was unusual in another respect. The alternative device concepts under consideration were developed to meet identical training requirements. Both concepts were intended to allow trainees practice on the same 42 preventive and corrective maintenance tasks (Koch, Pate, Pine, Herringa, & Carleton, 1979). Since the concepts did not differ with

respect to tasks performed on the devices, task commonality analysis was not appropriate in the present study.

Training objectives, task, and subtask lists were obtained from generic task analyses performed with the assistance of personnel at two Navy Class A electronics schools (Koch et al., 1979). These data provided the basis of the training effectiveness analysis. Three rating scales were used to compare the relative effectiveness of the 2-D and 3-D concepts: Similarity, Task Difficulty, and Task Criticality. Each of these is described below.

Similarity

Each subtask was assessed in light of preferred mode of simulation: 2-D or 3-D. In assessing 2-D vs 3-D modes of simulation, 2-D was judged to be preferred for all those task elements which did not demand dynamic interaction of controls and displays. The advantage of 2-D in such tasks, it was reasoned, stems from the fact that a 2-D display (eg., microfishe) could enhance the generalizability of generic tasks by displaying appropriate diagrams and pictures of various specific equipment to which the generic tasks apply. Such displays, it was reasoned, would extend the transfer of training. 3-D simulation was judged to be preferred for those tasks demanding dynamic interaction of controls and displays. A "1" was given to a trainer concept for each subtask on which it was preferred. Within a task, the number of preferences were summed to yield a 2-D similarity score and a 3-D similarity score.

Task Difficulty

Time to perform tasks was used to estimate learning difficulty. Estimated task completion times covered a span of .25 hrs. for the shortest tasks to 2.5 hrs. for the longest tasks. These times were then converted into a task difficulty rating scale as follows: 2 hrs. or greater = 3, 1 hr. to 2 hrs. = 2, and less than 1 hr. = 1.

Task Criticality

Subject matter experts rank ordered the tasks for criticality, i.e., relative importance of each task to be trained. Tasks were rank ordered from 1 to 42 with ranking #1 being of most importance. High reliability was obtained across all three judges. The rankings were then converted to a three point scale as follows: Tasks ranked 1-14 = 3, tasks 15-28 = 2, and tasks 29-42 = 1. This scale replaced the "Learning Deficit" scale in the TRAINVICE procedure because the judges could not differentiate the tasks on that scale.

The following formula was used to assess the training effectiveness of the 2-D and 3-D trainer concepts:

$$\text{Effectiveness} = \frac{\Sigma[(\overline{X}_{cd}) \times (\%S)]}{\Sigma \overline{X}_{cd}}$$

"C" is the criticality score expressed in terms of percent of the maximum criticality ranking obtainable. "D" represents the difficulty score expressed in terms of percent of the maximum difficulty ranking obtainable. "X_{cd}" represents an average of these two (C and D) scores. "%S" represents the percent of the maximum similarity score obtainable. The formula was applied to the rating scores for the 2-D and 3-D device concepts to yield an effectiveness score for each concept.

Results and Discussion

The effectiveness scores for the 2-D and 3-D concepts were .73 and .26, respectively, indicating a prediction of greater training effectiveness of the 2-D device. The training effectiveness scores obtained by applying TRAINVICE are seen to be related to extent of expected transfer of training to operational equipment (Wheaton, 1976b).

In the present context, the notion of transfer of training requires special consideration. For an equipment-specific trainer, transfer refers to the fact that skills acquired on the trainer are related to performance on a specific piece of equipment. If the trainer were effective, one would expect improved performance on specific operational equipment. Use of generic trainers requires a rethinking of the transfer of training concept. It would seem that the benefits derived from using a generic trainer would not necessarily be seen in performance on a specific piece of operational equipment. Rather, improvements in performance should be seen on a variety of operational equipment. For a generic trainer, the notion of generalizability of skills should be included in the concept of transfer of training.

In the present analysis of 2-D and 3-D trainer concepts, the notion of generalizability was incorporated into similarity comparisons. This consideration had the effect of making the 3-D trainer concept less desirable than the 2-D concept in many instances. Had generalizability not been a consideration, the effectiveness score for the 3-D concept would have been higher and that for the 2-D concept would have been lower.

It remains to be demonstrated that generalizability is a valid concept. However, some such concept appears to be needed in order to fairly judge the effectiveness of generic trainers. Empirical support for the use of such a concept is warranted in light of increasing interest in generic trainers and the subsequent need for tools to evaluate their effectiveness.

References

Hirshfeld, S., & Kochevar, J. Training device requirements documents guide (Honeywell Report 78 SRC 89). Orlando, Florida: Project Manager for Training Devices, Naval Training Equipment Center, January 1979.

Narva, M. A. Development of a systematic methodology for the application of judgmental data to the assessment of training device concepts (ARI Research Memorandum 79-7). Alexandria, Va.: U.S. Army Research Institute for the Behavioral and Social Sciences, May 1979.

Rose, A., Wheaton, G., Leonard, R., Jr., Fingerman, P., & Boycan, G. Evaluation of two tank gunnery trainers (ARI Research Memorandum 76-19), Alexandria, Va.: U.S. Army Research Institute for the Behavioral and Social Sciences, August 1976.

Wheaton, G., Rose, A., Fingerman, P., Leonard, R., Jr., & Boycan, G. Evaluation of three burst-on-target trainers (ARI Research Memorandum 76-18). Alexandria, Va.: U.S. Army Research Institute for the Behavioral and Social Sciences, August 1976. (a)

Wheaton, G., Rose, A., Fingerman, P., Leonard, R., Jr., & Boycan, G. Evaluation of the effectiveness of training devices: Validation of the predictive model (ARI Technical Report TR-76-A2). Alexandria, Va.: U.S. Army Research Institute for the Behavioral and Social Sciences, October 1976. (b)

Reference Notes

Note 1. Daniels, R. W., Heeringa, L. M., Koch, C. G., & Pine, S. M. Conceptual design for an EEMT simulator: Design alternative E (Report No. F2210-4.5). San Diego, Calif.: Navy Personnel Research and Development Center, April 1979.

Note 2. DoD Directive 5000.1. Major system acquisitions, Washington, D.C.: Department of Defense, January 1977.

Note 3. DoD Directive 5000.2. Major system acquisition process, Washington, D.C.: Department of Defense, January 1977.

Note 4. Heeringa, L., Koch, C., Persons, W., Daniels, R., Pine, S., & Squires, C. System specification for the three-dimensional unit of the electronic equipment maintenance training (EEMT) system (Report No. F2210-5). San Diego, Calif.: Navy Personnel Research and Development Center, June 1979.

Note 5. Koch, C., Pate, C., Pine, S., Heeringa, L., & Carleton, J. System definition for electronic equipment maintenance training system (Report No. F2210-2), San Diego, Calif.: Navy Personnel Research and Development Center, August 1979.

Formulation and Evaluation of a Methodology for Predicting Required Hands-On Training Following Simulator Training

William R. Bickley
U.S. Army Research Institute Field Unit
Fort Rucker, Alabama 36362

Abstract

Consider the function relating X, the amount of simulation training received, with Y, the amount of subsequent hands-on training required to attain a specified criterion. When X = 0, Y will be equivalent to the training required without simulation. As X increases, Y decreases rapidly at first, and then more slowly to some asymptote which is conceptually equivalent to non-simulable training that must be done hands-on. If the function is assumed to be of the form $Y = ae^{-b} \cdot + c$, then Ss may be administered various levels of X, their corresponding values of Y observed, and the three parameters determined by regression analysis techniques. The resulting regression equation can be used in determining cost and training effective mixes of simulator and hands-on training. This methodology was evaluated in the U.S. Army's acceptance tests of the AH-1 helicopter simulator.

Implicit in the armed services' development of any simulation training system is the assumption that training objectives are more economically attained through a mix of simulation and hands-on training than through hands-on training alone. The economy achieved is, of course, determined by the cost and the effectiveness of a unit of simulator training relative to the cost and the effectiveness of a unit of hands-on training. With past simulators, the cost differential between simulator and hands-on training has been so great that a marginally effective simulator might be used to realize overall training savings. However, as simulators have grown more complex and expensive to operate, the differential has shrunk to the point that precise quantitative determination of training effectiveness is becoming a major step in simulator testing and acceptance procedures.

Although this training effectiveness can be defined in many ways, the training psychologist's definition in this case is phrased in terms of trade-offs of simulator training for hands-on training: if the student receives x amount of relatively inexpensive training by simulation, y amount of additional relatively expensive hands-on training is then required to attain the training objective. To the training psychologist falls the tasks of (1) finding a suitable function relating units of simulator training with units of subsequently required hands-on training, (2) applying to the function the cost factors of the two training media, and then (3) minimizing the resulting cost function.

PRECEDING PAGE BLANK-NOT FILMED

The balance of this paper is in three parts. The first describes the derivation of an experimental methodology for finding the function just mentioned; the second describes the results of applying this methodology in the Army's operational evaluation of the AH-1 helicopter simulator; and the third describes how the data obtained can be used in the cost and training effectiveness analysis of a simulator. Although the example used is that of the AH-1 helicopter simulator, the methodology generalizes to tank simulators, to ship simulators, or to any training device used for initial skill acquisition.

Let us consider the function relating x, the amount of simulator training received, with y, the amount of subsequent hands-on training required to attain the training objectives. Intuitively, this function should have several characteristics. In general, it should be a decreasing function; that is, as amount of simulator training increases, amount of subsequent required hands-on training should decrease. However, the function should not be linear; the rate of decrease of y with increasing x should itself decrease until some asymptote is reached. The value of y at this asymptote could be conceived of as representative of those task elements which cannot be trained by simulation and require hands-on training. The y-intercept of the function corresponds, of course, to the situation in which training is totally hands-on and the simulator is not used.

Several general mathematical functions might be considered as candidates for characterizing the curve just described. But, as a practical measure, the function that will be considered is the general negative exponential, $y = ae^{-bx} + c$. This function is chosen for three reasons: first, it both fits the intuitive model just presented and is supported by what little empirical data exist in this area (e.g., Povenmire & Roscoe, 1973); second, similar functions have a long history in the development of quantitative models of incremental learning; and, third, it is mathematically tractable.

This model was developed in detail and then evaluated in the Army's OT II of the AH-1 helicopter flight and weapons systems simulator as follows.

METHOD

Subjects. Thirty-five aviators from regular AH-1 aircraft qualification classes at the U.S. Army Aviation Center (USAAVNC) participated in the study. All were qualified helicopter pilots transitioning to the AH-1.

Apparatus. The AH-1 flight simulator (AH1FS) is a high-technology training device which simulates the aircraft cockpit and instrumentation, aircraft motion and vibration, aircraft power plant and weapons noise, and out-the-window view. It is designed to afford training in visual contact flight, instrument flight, and weapons delivery techniques.

Procedure. The element of analysis chosen was the individual maneuver. Fourteen Ss received all training on all maneuvers in the AH-1 aircraft. The remaining 21 Ss received AH1FS training before proceeding to the aircraft. For each flight maneuver, each of these 21 Ss received one of 3 pre-specified amounts of training in the AH1FS. Data collected included number of attempts at each maneuver and amount of training time per maneuver attempt in both the simulator and the aircraft.

RESULTS

The data for each maneuver were fitted (using a least squares criterion) to the general function $y = ae^{-bx} + c$. For all maneuvers evaluated, the AH1FS was found to be an effective training device. Figure 1 shows example results for the maneuver "normal approach."

DISCUSSION

With a family of curves such as these the training psychologist has at his disposal the means of devising simulator-plus-aircraft curricula that approach optimization of cost and training effectiveness. Note that for any mix of simulator and aircraft training, total cost, C_T, can be expressed as the sum of simulator and aircraft costs, or $C_T = C_{sim}(x) + C_{a/c}(ae^{-bx} + c)$, which is shown graphically in Figure 2. As the figure suggests, this total cost function does have a minimum; mathematically, it is found by solving for the value of x for which the function's first derivative is zero. This technique was used in the cost and training effectiveness analysis of the AH1FS and the optimal training values found are to be utilized in the USAAVNC development of an AH-1 transition course incorporating the simulator.

So, at the practical level, the methodology has been demonstrated to be viable and of utility. As indicated previously, it can be applied to any training situation which involves simulator training followed by hands-on training. The nature of the model requires that at least 3 levels of simulator training be administered; however, there is no requirement that some Ss be used as a control group receiving hands-on training only, as is the case with most present transfer of training models.

At present, the model is being further developed to include additional predictor variables, such as aviator aptitude and experience level.

Reference

Povenmire, H.D., and Roscoe, S.N. Incremental transfer effectiveness of a ground-based general aviation trainer. Human Factors, 1973, 15, 534-542.

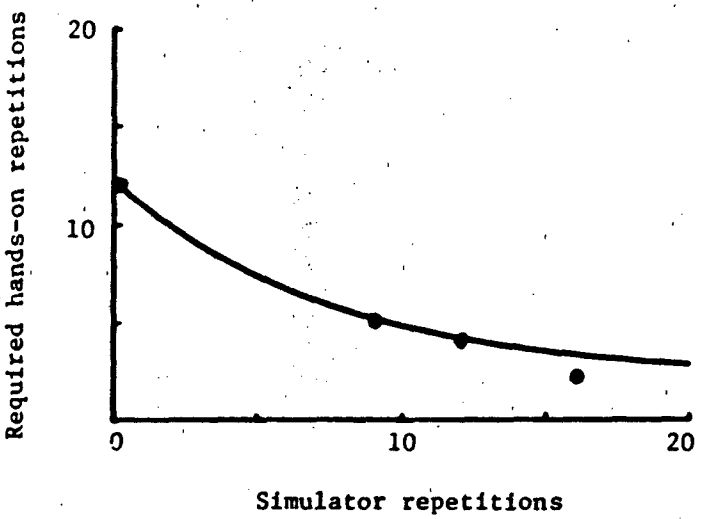


Figure 1. Simulator/hands-on tradeoff for "Normal Approach".

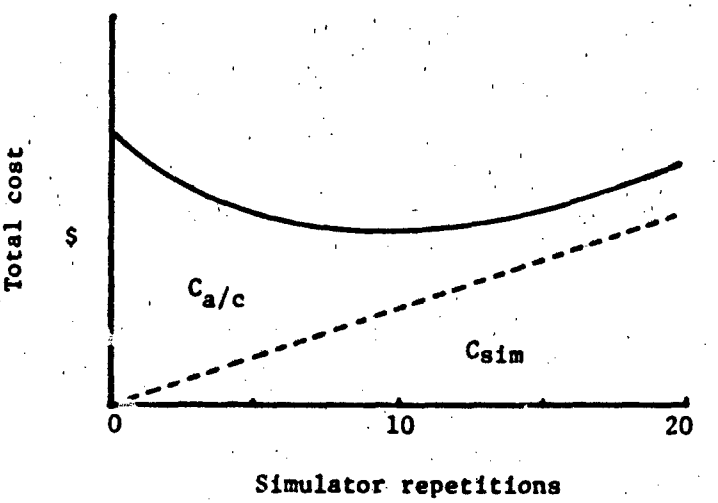


Figure 2. Total cost for simulator/hands-on mixes.

Simulation Training in the "Real World(s)": Some Issues & Empirical Answers!

Gerard M. Deignan, Air Force Human Resources Laboratory
Louis F. Cicchinelli, Denver Research Institute

Abstract

"Do simulators train students as effectively and efficiently as do actual hardware?" "How effectively do simulator and actual hardware trained students transfer skills to the "real" world of troubleshooting problems?"

To provide competent maintenance technicians to the field, expensive actual equipment is customarily employed for training. As a training device, actual equipment does not readily permit the controlled presentation of malfunctions representative of troubleshooting problems occurring frequently in operational settings. Less expensive real-time simulators do possess the capability for troubleshooting training, which incorporates hands-on practice to increase troubleshooting skill on samples of field-related maintenance problems. In addition to improved skills training, properly designed computer-based training simulators also have the potential to release more expensive actual equipment for field use. Major objectives of this evaluation were to compare two maintenance training simulators (3D & 2D) and actual maintenance equipment on dimensions of (1) instructional effectiveness, (2) cost-benefit, and (3) attitudinal acceptance within an operational ATC course.

The purpose of simulation is to provide learners with representative samples of actual tasks stripped of hazards to learners and equipment. In addition to an objective performance evaluation, properly designed interactive simulators provide the learner with direct, not vicarious task experience. To master tasks, learners must close the gap between their present state of knowledge, motivation, and related cognitive skills and that required for successful task performance.

Simulators in comparison to so-called "real systems" provide the opportunity to deliver instruction which permits greater

- Actual Task Training
- Skills Practice
- Task Specific Remediation
- Learner Diagnostics
- Motivational Consequences
- Objective Evaluation
- Instructor Control (Task & Learner)
- Cost-Effective Training

One serious limitation of real systems training is the immediate or delayed real consequences of inappropriate trainee responses. Real systems fidelity and complexity may also distract learners as they attempt to learn the subtle differences which define different concepts, processes or malfunctions. Moreover, the elapsed time to obtain feedback about the effects of troubleshooting strategies upon actual systems is often lengthy and in some cases several hours later. Due to hazards, equipment unreliability and limited training time, hands-on training in real systems is

generally confined to procedural tasks. Hence, trainees do not experience the range, nor complexity of real world troubleshooting tasks likely to be encountered on the job. To contrast, simulation presents the opportunity to reduce the job performance-training gap by providing experience with simulated tasks designed to improve (1) correct detection and diagnoses of simulated malfunctions and (2) application of appropriate knowledges and skills to correct malfunctions in less time than with actual equipment.

Despite the use of simulators for years, methodologically sound comparative studies of the instructional and cost effectiveness of simulators and actual equipment are conspicuously rare. Comparative simulator studies have been confounded with training content changes, lacked equivalence with baseline or control groups, or used questionable, subjective estimates of training effectiveness, time, or cost data as comparison criteria. These and related uncontrolled factors permit alternative explanations for obtained training and cost differences (Caro, 1977, Orlansky, 1977). The general finding of no skills differences as a function of simulator vs. actual equipment training is still, therefore, widespread. Additionally, few studies provide empirical data on transfer of training, incremental transfer effectiveness ratios or operating cost ratios (Orlansky, 1977). Differences in training effectiveness do not ensure comparable, or even any, cost savings. Similarly, cost differences may not indicate equivalent training effectiveness, nor should it be assumed. Rather, cost-benefit instead of cost-effectiveness ratios should be required of simulator vs. actual equipment comparisons.

Whether actual tasks are embedded in equipment with full or partial fidelity or confronted in the many different real worlds of operations, OJT or resident training, our knowledge of simulator and real equipment effectiveness for efficiently developing competent task performers, or evaluating them, is limited. The major objective of the present study is to determine the relative effectiveness of the 6883 actual equipment test station and the Honeywell and Burtek computer-based simulators on dimensions of:

- Instructional effectiveness
- Time-savings
- Cost elements
- Attitudinal acceptance
- Field performance

Method

Students. To gather preliminary and main experiments data, 115 male and female students scheduled for 2 1/2 days of practical exercises training within the 326X Course were used. An experimental (simulator trained) and control group (actual equipment trained students) comparison within the 6883 portion of the course was selected.

Procedure. Students were assigned randomly to the actual 6883 Convertor Flight Control Test Station or simulator (Honeywell). Comparison of the Burtek simulator with actual equipment is scheduled later in the year. To obtain an objective measure of instructional effectiveness, independent of the customary instructor subjective ratings of student performance within practical exercises, a troubleshooting performance test was developed with the aid of 326 personnel from three Air Force

bases. The troubleshooting performance test (TPT) constitutes the primary criterion by which instructional effectiveness comparisons are made. The TPT consists of a series of problem steps for isolating the actual malfunction from incorrect, more apparent alternatives. The task was presented on both the actual equipment and simulator. Performance measures recorded by an independent observer included total score, total time to completion, and sub-scores for specific sections of the test. To control for TPT test mode differences (i.e., tested on actual vs. simulator equipment), students were assigned to treatment, test mode, and followed-up for field assignment as depicted in Figure 1. To assess equivalency between simulator and actual equipment trained stduents, ASVAB and previous block scores were routinely gathered on each student for subsequent statistical score adjustments. Students subsequently completed interviews and attitude

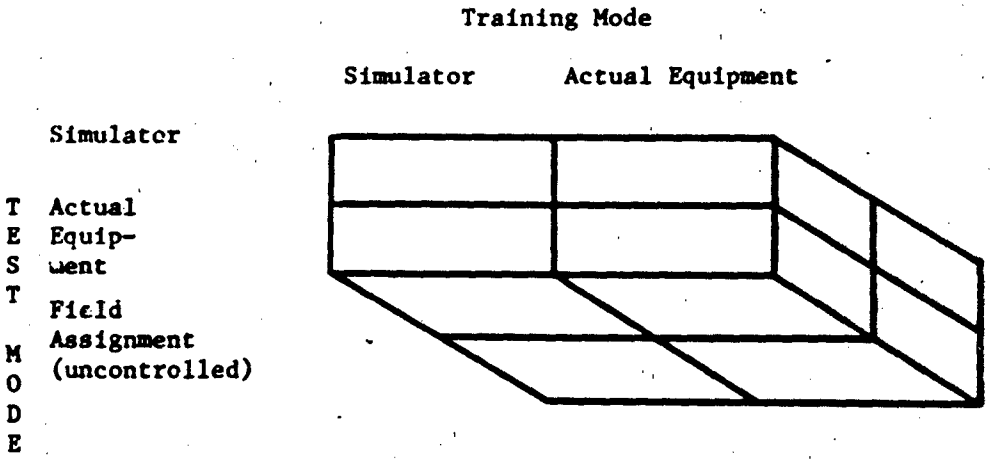


Figure 1. Experimental Design

questionnaires. Student follow-up of treatments effects included (a) performance in subsequent blocks of non-6883 training, (b) estimated job proficiency ratings of graduates by field supervisor, and (c) a paper and pencil test of understanding of concepts, procedures, and troubleshooting techniques. Development of an equivalent form of the troubleshooting test is currently in progress for field implementation. Comparative acquisition and remainder life cycle costs for both simulator and actual equipment were gathered by the contractor for cost model analyses. The cost model is a matrix of two primary dimensions composed of six major cost categories and three major life cycle phases. The basic cost model permits the development of scenarios in which cost category parameter values, e.g. student flow, aptitude level or course length may be varied to determine the impact upon cost-benefit ratios between actual equipment and simulator conditions.

Results

Data reported herein are essentially preliminary due to the major impact of administrative changes to airmen specialties during evaluation which in turn required former test station operators to be task proficient

in maintenance, i.e. troubleshooting skills. Course content, STS, and evaluation instruments were therefore adjusted to reflect the tasks of the new job specialty. In addition, periodic course revisions resulted in variable amounts of training times for different students. Thus, the following pilot data are based upon a small sample of students (N=57 simulator and N=58 actual 6883 trained students) exposed to (a) limited simulator or actual 6883 training (2 1/2 days of training within an approximately 18-week course) and (b) variable amounts of 6883 training due to course, STS, and instructional objectives revisions. Given future course stability, these student data would not be included in final summative evaluation results due to the obvious treatment confoundings. Reliable treatment differences at this stage are therefore not expected.

Instructional Effectiveness. No statistically significant total score achievement differences were obtained between groups on the troubleshooting performance test. Examination of ASVAB area scores (General, Administrative Mechanical and Selector Electronics) has indicated no aptitude differences among the four treatment groups shown in Figure 1. Similarly, no achievement differences were obtained between groups on the previous nine blocks of instruction.

Troubleshooting Time. Analysis of variance of troubleshooting time for treatments and test modes revealed significant effects for test mode $F(1,111)=9.31$, $p < .003$. No significant time differences were obtained between actual and simulator trained students. Simulator tested students took longer than students tested on actual equipment. This is due to a software-equipment problem within the simulator which is being corrected.

Cost Comparisons. Table 1 reports life cycle comparative cost data for actual equipment and the Honeywell simulator.

Table 1. Comparative Cost Data

Constant Cost Basis	Simulator	Actual Equipment
Acquisition and startup	$585,050	$1,965,859
Future, non-operating	77,800	77,800
Operating costs, 15 yrs	927,150	1,100,250
Total cost--1978 dollars	$1,590,000	$3,143,909

Discounted Cost Basis	Simulator	Actual Equipment
Acquisition and startup	$585,050	$1,965,859
Future costs (discounted at 10% over 15-yr life)	508,468	596,243
Total cost--present value	$1,093,518	$2,562,102

Summary: Comparative data are currently being gathered on the Honeywell 3D simulator and actual equipment. Comparison with the 2D Burtek simulator is expected in July 1980. Though preliminary data indicate no significant differences in achievement, this is due likely to: (1) inadequate training time (2 1/2 days), (2) technical order procedures emphasis as opposed to manual troubleshooting, (3) previous training transfer results in ceiling effects, and (4) insensitive/contaminated measure of objective criterion performance. Efforts are underway to develop objective, equivalent measures of performance and adequate training time extension. Life cycle cost comparisons reveal the 3D simulator is 58% less costly than the actual equipment. Hence, if simulators were found to be as, or more, instructionally effective than actual equipment, then generic simulators might profitably be deployed throughout entire maintenance training courses.

References

Caro, P.W. Some factors influencing Air Force simulator training effectiveness. Technical Report 77-2, Human Resources Research Organization, Alexandria, VA, 1977

Orlansky, J. Cost effectiveness of flight simulators for military training. Science and Technology Division Paper, Institute for Defense Analysis, Alexandria, VA, 1977

Design Requirements for A Planning Information System for Training and Personnel Technology RDT&E

Robert W. Stephenson
Air Force Human Resources Laboratory
Brooks Air Force Base, Texas 78235
and
Matilde K. Stephenson
St. Mary's University
San Antonio, Texas 78284

Abstract

Objectives and information requirements are proposed for the design of a planning information system for Training and Personnel Technology RDT&E. Objectives are (1) improved procedures for relating Training and Personnel Technology RDT&E to defense objectives; (2) continuous coordination of plans through the establishment of a computer-assisted, DoD-wide RDT&E plans information sharing system; (3) improved labeling of work units and simplification of top management presentations by organizing large numbers of small technically-described work units into broad thrusts that identify deliverables with obvious payoff for defense; and (4) more systematic procedures for program evaluation and resource allocation decisions. Issues are identified that should be addressed during the design stage of such a system.

The Department of Defense, like other organizations that conduct Research, Development, Test, and Evaluation (RDT&E), must evaluate plans, develop investment strategies, and establish information systems in order to make decisions about the allocation of resources. In the past, these activities have tended to be unstandardized between Services and somewhat informal. In recent years, formal procedures for documenting and reviewing Training and Personnel Technology RDT&E efforts have been developed, and people-related RDT&E has been coordinated at higher levels within the Office of the Secretary of Defense (OSD).

Difficulties in Marketing and Evaluating Plans for Training and Personnel Technology RDT&E.

Training and Personnel Technology RDT&E personnel often encounter difficulties in the marketing and evaluation of their plans. The plans are usually "softer" than hardware RDT&E plans, often take longer than expected, and don't seem to "fit" with the broader needs of the Department of Defense as well as they should. Many people also think that behavioral scientists are not as skillful at marketing their plans as they should be. The development of a planning information system, consequently, requires some advance thought about problems and requirements.

Part of the problem is that Training and Personnel Technology RDT&E really *is* different. For example:

(1) Most people-related RDT&E is conducted in order to meet the broader needs of the Services (e.g., improved productivity or a better person-job-match system). These broader needs may or may not support specific hardware systems, which tend to dominate the RDT&E

budgets of all the Services. It is difficult to incorporate people-related RDT&E into prioritization schemes that are primarily intended for decisions about hardware procurements.

(2) Support for people-related RDT&E has also been influenced by the fact that many of these efforts seem to benefit hundreds of activities without benefiting any one of them to a substantial degree. As a result, many recipients of these RDT&E products consider them "nice to have" rather than essential. Yet customer support is essential if people-related RDT&E is to have a beneficial impact on Service operations.

(3) Another consideration is that most people-related RDT&E tends to be efficiency-oriented rather than combat-survival oriented. At a time when four-day wars are seriously contemplated as a possibility, efficiency-oriented RDT&E often seems less necessary than equipment-oriented expenditures, such as the purchase of another tank, aircraft, or missile. Some military analysts hold that it is better to have inefficient tank operators with more tanks than efficient operators with fewer tanks. There must be a function that states optimum relations, since it would be equally wrong at the limits to have expert operators with no tanks or many tanks with no trained operators.

The Need for a New Planning Information System.

A new planning information system could help resolve many current difficulties in the evaluation and marketing of plans for Training and Personnel Technology RDT&E. This is a good time to develop a system of this type. As noted by Taylor, many Congressional committees and their staffs have been concerned with how R&D products have been used to improve military capability. New planning systems and procedures might help alleviate these concerns. Another issue is duplication of effort due to a perceived lack of managerial and technical coordination and cooperation (Taylor, 1979). Information systems that can be shared by researchers and managers in many different parts of the country are now feasible with current computer capabilities.

Previous Work on the Design of RDT&E Management Systems.

Several reviews of the literature on RDT&E plans evaluation, project selection, and resource allocation have been published (e.g., Baker and Pound, 1964; Cetron and Martino, 1967; Augood, 1973; Clarke, 1974; and Souder, 1978). References of special interest are Nutt (1965) and Schoman, Dick, and McNight (1969), both of which describe investment models specifically designed for Department of Defense (DoD) agencies.

When Souder wrote his review in 1978, many years had elapsed since work on project selection models was initiated. From this historical perspective, Souder concludes that most project selection and evaluation models have not been useful because they were not designed to deal with real world environments. For example, the typical management science model seems to assume a single decision maker in a well behaved environment with accurate information. The real world, in contrast, usually involves many decision makers and many decision influencers in a dynamic organizational environment with information that is anything but perfect. Management science models also seem to assume that the goals of RDT&E are well known and invariant, whereas the real world usually involves ever changing, fuzzy goals, with the information needed to evaluate these goals scattered throughout the organization. People-related

RDT&E organizations are also characterized by conflicting multiple objectives and multiple constraints, some of which are not economic in nature. In short, the real world is not as simple, logical, or quantifiable as management scientists would like it to be.

Major Design Objectives.

Despite a complex situation with unusually difficult design problems, a DoD-wide, plans-oriented management information system for Training and Personnel Technology RDT&E is possible. The following design objectives for a system of this type are proposed:

(1) Improved procedures for relating Training and Personnel Technology RDT&E to defense objectives. In spite of one admirable attempt to do so by the staff of the Assistant Secretary of Defense for Manpower, Reserve Affairs, and Logistics (Department of Defense, 1979), plans for people-related RDT&E have not been satisfactorily related to short and long range military defense objectives as seen by higher echelons. Some changes in the objectives or the way in which defense objectives are seen by higher echelons may be needed. A plans evaluation system that can give more recognition to diffuse contributions of multipurpose Training and Personnel Technology RDT&E is also needed.

(2) Continuous coordination of plans through the establishment of a computer-assisted, DoD-wide, RDT&E plans information sharing system. It is difficult to coordinate plans. They are always changing, and they are typically not very well defined. Difficult or not, however, coordination of plans in DoD is absolutely essential. The commanders of the various Training and Personnel Technology RDT&E organizations have been very active in this area, largely by scheduling multilevel coordination conferences (e.g., Sands, 1979). More work is needed in this area, however, and additional work on interservice coordination is an important design consideration for a new system.

A computer-assisted DoD-wide plans information sharing system seems the best way to facilitate interservice coordination and reassure congressional committees and their staffs that there is no duplication of effort. Current systems of this type (e.g., Management and Scientific Information System (MASIS) and Defense Documentation Center searches) are largely limited to periodic checks of completed or on-going work. In contrast, a plans-oriented system could be used to provide a continuous check on future plans; to revise budgets for future years; to minimize planning paperwork by providing a flexibly updated record system; and to store the input data needed for program evaluation and investment strategy decisions.

(3) Improved labeling of work units and simplification of top management presentations. As noted by Dr. Ruth M. Davis, Deputy Under Secretary of Defense for Research and Engineering, "In some quarters of the Armed Services Committee, there is a feeling that certain projects are a total waste and that nothing of operational value is coming from people-related R&D programs " (Sands, 1979, p. 5). One reason for this lack of enthusiasm for people-related RDT&E is the way in which the multiplicity of RDT&E efforts is organized and labeled.

There are more than a thousand people-related RDT&E efforts in the Department of Defense, and the titles usually specify technical objectives without any clear indication as to what the deliverable items

may be. Higher level managers often look at titles without obtaining details or determining the contribution of the effort to defense objectives. They consequently tend to think that the many diverse efforts with technical titles are not making important contributions. If these hundreds of titles were organized into a few broad thrust headings that identify clear-cut products with obvious payoff for defense, communications would be greatly improved. It should be noted, however, that this requires changes in the way that RDT&E is managed--not just the way in which it is labeled. Small isolated RDT&E efforts that are not part of a larger thrust (and cannot be justified as basic research) must be scrutinized closely to be certain that they really deserve priority and do not have titles that will be seen as irrelevant for important defense objectives. The Air Force Human Resources Laboratory is currently implementing a thrust management approach of this type.

(4) More systematic procedures for program evaluation and resource allocation decisions. Granted that most program evaluation and investment strategy decision methods make assumptions that do not correspond with the real world, there are still many techniques and procedures that could be useful. For example, a new planning information system for Training and Personnel Technology RDT&E should provide managers with information that would help them to reallocate resources if the Laboratory budget were to be substantially increased or decreased. A system could be designed to do this. Ideally, the information system would also consider the "missed opportunity costs" of each set of allocation decisions, by comparing the merits of efforts that are not funded with those that are. Any system that does not try to incorporate some sort of systematic resource allocation subsystem is not taking advantage of the state of the art.

Design Issues

A number of systems design issues have been suggested in the literature. Sigford and Parvin (1965) talk about 12 "mission concepts"; Detron, Martino, and Roepcke (1967) identify 15 "features" of methods for evaluating and selecting R&D projects; Krasnican (1971) identifies 12 criteria for planning system evaluation; and Clarke (1974) identifies 8 basic questions about the innovation decision process. Reference is also made to comments about information requirements attributed to Dr. Davis by Sand (1979) and Taylor (1979). Drawing generally from these references, the authors developed a list of issues that need to be dealt with during the design stage of a new system. The design issues are organized into five categories as follows:

(1) How should plans be evaluated? Procedures are needed for: (a) deciding which factors should be considered at each decision point; (b) restating defense objectives so that they can be more easily translated into requirements for people-related RDT&E; (c) scoping proposals to a reasonable period of time and insuring adequate use of previously available data or knowledge; (d) identifying overlapping and redundant projects; (e) collecting evaluative information at specified points in time; (f) deciding if, when, and how mathematical models should be used for basic research, exploratory development, and advanced development; and (g) tracking possible changes in the nature of the problem and possible progress in closely related RDT&E.

(2) How should resources be allocated? Plans are needed for: (a) deciding what kinds of resource allocation decisions should be made, who should make the decisions, and when they should be made; (b) rapidly reallocating funds when budget cuts take place or additional funds are made available on short notice; (c) determining the cost consequences of a change in priorities; and (d) considering "missed opportunity" costs.

(3) What kind of an information system is needed? Detailed plans are needed for: (a) inputs, outputs, and information flows; (b) the amount and kind of computer assistance to be provided at each point; (c) the way in which input data are coded, stored, processed, and retrieved; (d) output options; (e) storage requirements; (f) equipment requirements; (g) ways of identifying and weighting systems evaluation criteria; and (h) the costs of various options with respect to inputs, outputs, hardware configurations, operation and maintenance costs, and noneconomic considerations.

(4) How could coordination be improved? Procedures are needed for: (a) analyzing interrelationships among projects in order to decide how much coordination is needed; (b) documenting technical as well as managerial coordination; and (c) identifying the conditions under which organizational specialization (e.g., "lead agency" status) would be desirable.

(5) What kinds of information should be provided to top management? The following options with respect to improved management information systems should be considered: (a) documenting RDT&E requirements by describing Service specific problems for which users or implementers are waiting for RDT&E results; (b) documenting past contributions by providing examples of how past products have been used or have provided improvements in capabilities; (c) illustrating expected contributions with easy to understand examples of policies or actions that could implement successful RDT&E; (d) grouping smaller efforts into large thrusts that have clearcut relevance for defense objectives; and (e) providing summary information about life cycle costs and projected benefits.

References

Augood, Derek R. A review of R&D evaluation methods. IEEE Transactions on Engineering Management, 1973, EM-20, 114-120.

Baker, N. R. and Pound, W. H. R and D project selection: where we stand. IEEE Transactions on Engineering Management, 1964, EM-11, 124-134.

Cetron, Marvin J., Martino, Joseph, and Roepcke, Lewis. The selection of R&D program content--survey of quantitative methods. IEEE Transactions on Engineering Management, 1967, EM-14, 4-12.

Clarke, Thomas E. Decision-making in technologically based organizations: a survey of present practice. IEEE Transactions on Engineering Management, 1974, EM-21, 9-23.

Department of Defense. Manpower and Logistics Research Agenda, June 1979.

Krasnican, Milan J. Managing the development of an experimental computer-aided technology planning system (PLANET). IEEE Transactions on Engineering Management, 1971, EM-18, 50-57.

Nutt, A. B. An approach to research and development effectiveness. IEEE Transactions on Engineering Management, 1965, EM-12, 103-112.

Sand, Franklin F. The 1979 interservice conference for the coordination of people-related R&D; a summary (Special Report 79-17). San Diego, Calif.: Naval Personnel Research and Development Center, April 1979.

Schoman, C. M., Jr., Dick, D. N., and McNight, T. R. Relating organization goals and technological forecasting for research and development resource allocation. IEEE Transactions on Engineering Management, 1969, EM-16, 148-161.

Sigford, J. V. and Parvin, R. H. Project PATTERN: a methodology for determining relevance in complex decision making. IEEE Transactions on Engineering Management, 1965, EM-12, 9-13.

Souder, William E. A system for using R&D project evaluation methods. Research Management, September 1978, 21, 29-37.

Taylor, Henry L. Politics of instructional innovation; program justification, development, and apolications. Paper presented at the American Psychological Association Convention, New York, September 2, 1979.

The Evolving Role of AMEDD Clinical Community Psychology: Utilization of the Clinical Community Specialist

LTC(P) Cecil B. Harris, Ph.D
School of Medicine
Uniformed Services University of the Health Sciences
Bethesda, MD 20015
and
Psychology Consultant
To The Surgeon General
Department of the Army
Washington, DC 20310

I have been asked to discuss the proposed utlization of graduates of our Community Psychology postdoctoral fellowship. Before addressing this topic directly I will attempt to provide a framework for our discussion about this much needed specialist.

This paper will briefly trace the development of clinical psychology in the Army and include aspects of the changing role of the profession. The focus will then shift to the military as a special environment for clinical practice and conclude with our current thinking on the proposed utilization for our training product.

Although psychologists have delivered services to the military since WWI, clinical psychology did not formally fill an identified role in the Army Medical Department (AMEDD) until 1945. Characteristic of the times, the military psychologist was limited to the role of test administration. Psychologists had played a similar role as members of the Adjutant General's Corps throughout the War and for some time before. Psychologists were heavily involved in screening illiterates and mentally retarded at our various induction centers. A few psychologists had functioned in a slightly broader role at the mental hygiene clinics which were established at the beginning of WWII. These psychologists conducted therapy and participated in medical and physical evaluation board actions.

The Surgeon General formally established clinical psychology in the AMEDD in September 1944. With the majority of psycholgists serving in the Medical Administrative Corps, the former name of the current Medical Service Corps, it was not long before all psychologists were transferred from The Adjutant General to The Surgeon General. It is significant that the doctorate has been the practice degree in AMEDD psychology throughout the years and reflects the commitment to parallel role development with the Veterans Administration and non-federal civilian practice.

Clinical psychology is an evolving profession. The expertise that the specialized training affords the clinical psychologist prepares this health service provider to deal directly with people who have problems requiring professional intervention. The traditional areas of psychodiagnostic testing and individual and group psychotherapy have been expanded to add new roles and services. Crisis intervention, brief therapies, family therapy and community mental health care all apart of the evolving new directions for clinical psychology. Concurrently, the search for new paradigms of practice has lead to a shift in emphasis from

secondary and tertiary prevention to primary prevention. This shift in focus has moved the practice of clinical psychology from an individual to community focus.

The community psychology movement focuses on prevention as opposed to treatment. Clinical psychologists who have broadened their skills with this new focus seeks to improve the quality of mental health for a defined population by working with communities, systems, and organizations which make up the subsystems of these communities. Points of intervention have shifted to planners, caretakers, social institutions and political organizations. To many this represented a revolution in behavioral sciences. Health providers were once again looking to the person-environment interactions when evaluating behavioral dysfunctions.

The ferment in behavioral science which had been evident for years lead to the May, 1965 conference at Swampscott, Massachusetts and the official birth of community psychology. The direction charted by the participants at this conference was given further support by the National Training Conference on Community Psychology held in Austin, Texas in 1975. Community psychology was given the mission of reducing the situational and social forces which contribute to dysfunctional behavior and supporting those forces which contribute to coping strengths and competences. Thus, individual clinical psychology practice shifted for many to clinical-community practice. This was not new for military psychologists, but merely represented the parallel civilian practice moving a step closer to the long recognized reality within the military environment. The Army Mental Health Consultation Service (MHCS), born in the post WWII years, is institutionalized recognition of the fact that much of behavior reflects the interaction between persons and their environments. In fact, the military environment is a special area which is ripe for clinical-community intervention at every level.

One can gain a clear view of the enormity of the military environment from the following statistics which were provided by Mr. Vernon McKenzie, Acting Assistant Secretary of Defense (Health Affairs) in testimony before the Military Compensation Subcommittee of the Committee on Armed Services, House of Representatives, Ninety-Sixth Congress. There are about 9 million eligible beneficiaries of the DOD health care system. Active military personnel number 2.1 million, active duty dependents 2.9 million, retirees 1.2 million, with 2.3 million dependents of retirees and 0.4 million survivors of members or former members included in the 9 million. Mr. McKenzie estimates that 75 percent of these beneficiaries live within a 40 mile catchment area of military hospitals. In terms of facilities the DOD operates 165 hospitals with 127 of these within the continental United States (CONUS). The Army operates 49 of these facilities, the Navy 35 and the Air Force 81. FY 1979 there was an estimated 83,900 military and 42,000 civilian supporting the DOD health delivery effort. Included in these manpower figures are 10,819 military physicians, 4,936 military dentists, 10,221 military nurses, 7,431 medical service officers and 73,342 enlisted personnel in the direct-care system. This is indeed a special environment for the practice of clinical-community psychology.

We have less than 100 clinical psychologists in the Army to serve

our share of this special environment. Many of our practitioners have not moved away from the traditional individual practice; others are forced to remain in that role due to the requirements of a manpower accountability system which supports their authorized position. A large number of our clinical psychologists have abandoned the challenge of this special environment. For them the challenge is too great, or their morale has suffered because they believe the environment has not reinforced them properly. The proposed community psychology fellowship is designed to change this sense of impotence for psychologists as well as broaden the providers' armamentarium for serving the special military environment.

Utilization of the clinical-community psychologist is virtually limitless within our special environment. Since this is utopian we must consider several strategic points of intervention. Early products of the fellowship can best be utilized in teaching roles at each of our four internship training sites. Clinical-community skills have been a core area of emphasis at each site for the past ten years. While the skills were emphasized there was not an expertly trained staff member to serve as a role model and direct the interns' development in these areas. Of equal importance will be the assignment of clinical-community psychologists to serve in staff positions at Department of Defense (DOD), Department of the Army (DA); Health Services Command (HSC) and other major command (MACOM) headquarters. These key action officer/advisor positions are crucial in ameliorating many person-environment problems which accrue from policy decisions and actions.

Since the training program is multi-disciplinary, behavioral scientists from each of the disciplines could soon have impact throughout the system. A true shift from secondary/tertiary to primary prevention could be readily achieved. Thus, at the microlevel our clinical-community psychologists can have early impact by multiplying their impact rapidly at the training sites. The graduates of the internships will be better prepared for their basic developmental (0-7 years) roles as change agents at combat division, Medical Department Activity (MEDDAC), and Community Mental Health Activity (CMHA) locations. Intermediate developmental (7-15 years) utilization at Medical Center (MEDCEN), service school and headquarters staffs will be facilitated by the direct assignment of mid career qualified clinical-community psychologists to positions of leadership at these locations. Clinical-community psychologists making their advanced (15-23 years) and major (23+ years) professional contributions can have wide influence serving in consultant roles at regional, HSC, MACOM, and DOD/DA levels.

The continued evolution of clinical psychology within the AMEDD is furthered by institutionalizing the expanding role through the community psychology fellowship. More importantly, service to the special military environment is greatly improved by this shift in emphasis. Conserving the fighting strength is far more effective at the primary level of intervention.

U.S. Army Residential Alcohol Treatment Facility: A Proposal

Harry D. Silsby, LTC, MC
Chief, Department of Psychiatry
William Beaumont Army Medical Center
El Paso, Texas 79920

Abstract

A proposed residential alcohol treatment facility (ATF) for U.S. Army is described. Pending final approval, this ATF will offer treatment in 1980 to chronic, career alcoholics. A multidisciplined staff of 17 will provide necessary health care and administrative services at William Beaumont Army Medical Center (WBAMC) during a six-week residential period. Families and spouse will be involved in treatment which can be accomplished at the local installation. An aggressive follow-up program will be instituted to insure feedback regarding patient progress.

Mission

The mission of the Army Residential Alcohol Treatment Facility will be to provide a comprehensive therapeutic program for chronic, career alcoholics and to provide appropriate follow-up. The Treatment Facility will also be responsible for education and orientation of medical staff and line commanders. The ATF will conduct ongoing research in the area of alcoholism and dependency disorders.

Rational and Discussion

For some time it has been apparent that the Army's Alcohol Drug Abuse Prevention and Control Program (ADAPCP) has lacked the capacity for long-term residential treatment for chronic career alcoholics. Residential programs of this sort have well documented the cost effectiveness of such treatment approaches, both in industry and the military, as exemplified by the U.S. Navy's program in Long Beach. In the past, Army officers and senior NCO's have been treated at the Navy's program, as well as the Tri-Service Facility at Bethesda, Maryland. These facilities, however, are currently maintaining lengthy waiting lists, and treatment for Army personnel is not readily available.

U.S. Army Health Services Command is presently proposing the establishment of a 30 bed treatment facility at WBAMC. Construction costs of this program are negligible in light of presently assigned facilities.

Treatment Method

A multidisciplinary treatment regimen will be provided, to include necessary medical care, psychiatric and psychological treatment, alcohol education, and other techniques such as individual counseling, group therapy, Alcoholics Anonymous (AA), antabuse, occupational therapy, physical therapy, family therapy, work therapy, milieu therapy, and career counseling. Other modalities such as TA/Gestalt, psychodrama, biofeedback, and condition reflex therapy would be available as indicated. Duration of treatment will be approximately six weeks which will not include time spent in detoxification at the local level or a Medical Treatment Facility for other medical complications.

PRECEDING PAGE BLANK-NOT FILMED

Contact will be maintained by the ATF counselor with the clinical director at the local installation.

The ATF will recommend appropriate follow-up treatment and this will be coordinated through the local ADAPCP clinical director, immediate supervisor and AA, who will monitor client progress. Family involvement in treatment will also continue at the local program level during the follow-up phase.

Transition and Follow-up: An aggressive and systematic method of follow-up will be instituted so that the ATF can receive feedback regarding progress of the ex-residents. This will include maintenance of certain statistics to determine success ratios of servicemembers who complete treatment.

Summary

This paper has described a new multidisciplined alcohol treatment program for U.S. Army personnel and their families. While this is a proposal, the program is confidently expected to begin within this fiscal year. Further information is available from the author upon request.

Substance Abuse in the Combat Environment
The Heroin Epidemic

Harry D. Silsby, LTC, MC
Chief, Department of Psychiatry
William Beaumont Army Medical Center
El Paso, Texas 79920

Abstract

From the existing literature, substance abuse among American Troops has been fairly consistent from location to location, with the major exception of heroin abuse in the early 70's in the Republic of Vietnam. It is the author's opinion that substance abuse in the military, for the most part, tends to parallel substance abuse in the civilian population with the exception of a few remote areas where substance accessibility exceeds that in the Continental United States. There was no doubt, however, that heroin abuse in the early 1970's in the Republic of Vietnam reached epidemic proportions. Of the approximately 50% heroin experimenters, only 10% were felt to be chronic users to the point of physiological dependence. There is little doubt that this latter statistic had an impact on combat effectiveness.

Introduction

Substance abuse among American youth has been a subject of great interest and concern since its identification in the mid-1960's. Of equal concern has been the military's role as an etiological agent for increased substance abuse in young troops, as well as the impact of drug abuse on combat effectiveness. Since the United States' involvement in the Vietnam Conflict, there has also been concern as to the use of heroin by servicemen stationed there. This paper will address these issues, as well as review the existing literature from the period.

Overall Incidence of Substance Abuse in the United States Army

Stanton[7] reporting in the Archives of General Psychiatry on a study conducted in Vietnam in 1969, reported the incidence of drug abuse in outgoing enlisted personnel was 50.1% use of marijuana, 16.2% use of amphetamines, and 11.6% use of heroin or morphine; however, opium use was reported as 17.4%, and hallucinogens were reported as 5.3%. He found that opium use among personnel ages 17-26 nearly tripled in Vietnam, according to the responses gathered from the outgoing enlisted personnel. These findings tended to be corroborated by the Sapol[4] study reported in the Journal of American Pharmaceutical Association.

Greden and Morgan[2] reported on drug use patterns in the military from Fort Lee Virginia, in 1970. They reported that 50.5% of the total population sampled reported at least some illicit drug use, most of which was presumably marijuana; and 10.6% reported "heavy" use (more than ten times weekly). Their demographic profile of the frequent drug user was a person in his teens, single, with less than a high school education, rank less than E3, and from a large city. A later study among troops in the

United States was reported by Callan and Patterson[1] in 1971. In their study, approximately six thousand inductees were surveyed with the following incidence of previous substance abuse experienced: marijuana 78%, amphetamines 38%, barbiturates 28%, heroin 8.59%, opium 13.25%, and LSD 29%. They further reported little racial difference in substance abuse patterns with the onset generally in the late teen years. They further felt that 79% of the substance abuse took place in a social setting or was used experimentally. About the same time period, from surveys in the United States Army in Europe, Tennant[8] reported from a survey of approximately five thousand troops, 46% had experienced some form of illicit substance use, with 16% using more than three times per week. Usage of opiate derivatives, however, was approximately 1.5% on a frequent basis. Silsby[6], in a study conducted in 1972, also in Europe, reported in the Journal of Military Medicine that 56% of the study population had used illicit substances, and 10% of the population used on a daily basis.

From these data, it can be concluded that the overall incidence of illicit substance usage in the young American soldier was somewhere slightly over 50%, most of which was experimental or social. The most frequently used substance was cannabis; and it can also be concluded that 3-10% of the population was using illicit substance on a daily basis. These estimates seem to be relatively consistent whether considering young inductees, soldiers stationed in Europe, or soldiers in Vietnam prior to 1970. They also tend to correlate with studies from the civilian population.

The "Heroin Epidemic" in the Republic of Vietnam

As pointed out above, the substance abuse incidence rates were fairly constant among young United States Army troops until 1970 in the Republic of Vietnam, at which time a marked increase in the use of heroin became apparent. Heroin, referred to by the troops as "skag," was supplied cheaply and readily in small plastic vials of approximately 200 milligrams of 98% pure heroin, which sold in country from three to five dollars per vial. Many of the troops were actually unaware of the nature of the substance initially. The main route of ingestion was either sniffing or smoking. Siegel[5], reporting from the United States Army Hospital in Bangkok, confirmed widespread use of heroin to epidemic proportions among United States Army troops stationed in Thailand, which seemed to closely parallel that seen in Vietnam. Robins, Davis, and Goodwin[3] reported in the American Journal of Epidemiology that almost half the Army enlisted men who left Vietnam in September 1971 had tried one or more narcotic drugs. They found, in a population of 943 returnees, that 43% had tried a narcotic substance, 34% heroin, and 38% opium. The majority of the users were Regular Army as opposed to draftees, had little education, came from broken homes, and large metropolitan areas. They further reported that almost every enlisted man was approached by someone offering heroin within the first month of his arrival. About one-fifth of those using narcotics used them with sufficient regularity to develop some signs of physiological dependence. Our findings from American troops stationed in Military Region IV in late 1970 and early 1971 tended to corroborate the Robins study. In that area, a survey run among both combat and support troops indicated 45% use or experimentation with heroin or opium.

Detoxification experience in Can Tho, Vietnam, from January until April 1970 recorded 184 troops admitted with heroin dependence and 165 completing a one-week detoxification program, at the end of which those successfully completing were considered drug free. Follow-up studies after three weeks, however, revealed that only 54% of those detoxified remained drug free.

Discussion

There is little doubt that substance abuse, starting in the mid-1960's was prevalent in the United States Army, as well as the civilian population The incidence, for the most part, tended to be fairly constant among groups from location to location, with slight variations probably accounted for by availability, cultural shock, and peer group pressure. The overall incidence rates can be predicted to have been somewhere over 50% of the population engaging in some form of experimentation or social illicit drug use, with less than 10% ever becoming involved to the point of physiological or psychological dependence. The major exception to this, however, was seen in the Republic of Vietnam. At this time, 96-98% pure heroin became readily available. Experimentation with this substance became the rule rather than the exception among lower ranking enlisted personnel.

A number of hypotheses have been put forward: (1) that the heroin was less easily detected and much more easily stored than marijuana, (2) that military authorities had lost credibility, especially in their handling of marijuana, and (3) social unrest and discontent regarding the war in Vietnam. The overall contributing factor, however, in the author's opinion, was the ready availability of the substance. During this time, a great deal of societal and political concern was generated regarding the long range outcome of the Vietnam veteran. In this regard, the article by Robins, Davis, and Goodwin[3] has been most encouraging. Of the 43% narcotics users in Vietnam, only 9.5% of the population reported any use since return from Vietnam; and only a .7% reported any physiological addiction. Of an equally large group of soldiers who were discharged from the Republic of Vietnam and picked up prior to discharge as narcotic positive on urinalyses, only 33% of those individuals reported using after return to the States, and only 7.2% reported physiological addiction. These data seem to point to the fact that when the Vietnam veteran was out of the area of high availability, usage tended to return to the predictable endemic levels.

The final, and perhaps most important, question from the military point of view is: "What effect do illicit substances have on combat effectiveness?" Stanton reports in his article that marijuana use does have some correlation with exposure to enemy fire, which was also confirmed by the Sapol study reporting greater marijuana usage in soldiers exposed to combat. Although some of the tranquilizing effects of cannabis may actually have been beneficial, there is no doubt but that there was some effect on combat effectiveness, but this was probably minimal. On the other hand, heroin can be more incapacitating, not only in acute effect but in terms of physiological dependence. Of the predicted 40-50% heroin experimenters, it is estimated that combat effectiveness was probably affected adversely by only the chronic, dependent users; but that number

was significant to the combat commander.

Bibliography

1. Callan, J.P., Patterson, C.D.: Patterns of Drug Abuse Among Military Inductees, Am J Psychiat, 130:3, 260-264, 1973.
2. Greden and Morgan, D.W.: Patterns of Drug Use and Attitudes Toward Treatment in a Military Population, Arch Gen Psychiat, 26:113-117, 1972.
3. Robins, L.N., Davis, D.H., Goodwin, D.: Drug Use by U.S. Army Enlisted Men in Vietnam: A Follow-up on their Return Home, Amer J of Epidemiology, 99:4, 235-249, 1974.
4. Sapol, E., Roffman, R.A.: Marihuana in Vietnam, J. Amer Pharm Assoc, 9:615-619, 1969.
5. Siegel, A.J.: The Heroin Crisis Among US Forces in Southeast Asia: An Overview, JAMA, 223:11, 1258-1261, 1973.
6. Silsby, H.D., Lawson, T.L, Hazelhurst, C.D.: Drug Abuse Prevention in the Military: A Punitive/Administrative Action Report, Military Medicine, 486-487, Jul 75.
7. Stanton: Drug Use in Vietnam, Arch Gen Psychiat, 26:279-286, 1972.
8. Tennant, F.S.: Drug Abuse in the US Army, Europe, JAMA, 221:10, 1146-1149, 1972.

Post-Graduate Fellowship in Community Mental Health: A Proposal

Gary R. Greenfield, CPT, MSC
Chief, Community Mental Health Activity
Jack E. Bentham, MAJ, MSC
Chief, Psychology Service
William Beaumont Army Medical Center
El Paso, Texas 79920

Abstract

A proposal for a post-graduate Fellowship in Community Mental Health (CMH) is described. Pending DA approval, the first Fellow is expected to begin a one year training program in September 1980. The training will be a combination of didactic and experiential work developed by William Beaumont Army Medical Center (WBAMC) and Texas Tech University (TTU) personnel, following general guidelines as suggested by the Division of Community Psychology, American Psychological Association. A certificate of post-graduate training will be awarded by TTU. The Fellowship will be initially restricted to Army behavioral scientists. Depending on demand, the program can be expanded to a tri-service level.

The following outline replicates in principle the necessary elements in designing a training program in Community Psychology (CP) suggested by the 1st Annual Training Conference in Community Psychology, sponsored by the Division of Community Psychology, American Psychological Association. Adherence to these guidelines is expected to facilitate eventual accreditation of the Fellowship. An earlier paper by Luchsinger, et al (1978) delineated the principles underlying the Fellowship and some implications for the DOD behavioral science professional.

Objectives

Training objectives seek as a minimum to develop assessment, intervention and evaluation skills at all levels of organizational and community hierarchies from small groups to large systems.

Specifically, the Fellowship objectives are to prepare the Fellow to: 1) direct and manage a CMH Center. Historically, behavioral scientists have not been trained in management, administration, and finance and budgetary matters. Collectively, these areas form a baseline of skills which are seen as essential to the graduate of the Fellowship. In addition to direct clinical skills, the Fellow will significantly add to his repertoire of skills which will increase his effectiveness and utility to the AMEDD; 2) function on a higher organizational level as a consultant to command at installations, MACOM, Service School and DA level. Service at these various levels will allow systematic input at other organizational levels reaching broader populations and thus different "communities." This not only provides direction for the training but it also begins to define possible career tracts for the graduate; 3) obtain the skills necessary to function at strategy level 2 (CMH) and 3 (CP) as described by Goodstein and Sandler (1978). Skills identified here appear to represent "the state-of-the-art" in CP.

Fellowship Format

The format for the post-doctoral program includes a completion of seminars, workshops, and applied field experience. The primary thrust of the program will be involvement in community field experience through specific agency/organization placements, with tutorial courses taken with selected professors at TTU, Lubbock, Texas, to be supplemented by workshops offered in modular fashion at WBAMC. The Fellow would complete a minimum of 15 credit hours during the Fellowship year (September thru the following October) and be awarded a certificate of post-graduate study by TTU. This maximizes field experience while insuring a minimum acceptable knowledge and theoretical base. Field placement objectives and tutorials will be individually planned by the Fellow and his adviser-advocate.

The Chief of Psychology Service, together with the Chief, Department of Psychiatry, Chief, Social Work Service, Chief, CMH Activity, Director of Fellowship training and a faculty representative from TTU will serve as the Fellowship Committee. Each Fellow will have a committee member designated as his adviser-advocate. This committee will be chaired by the Director of Fellowship training who will be a military psychologist with specialty training in CMH and CP. The Fellow would be expected to meet monthly with each committee member. Nationally known academicians in the community arena will be invited to sit as external consultants to the Fellowship Committee.

Quarterly progress conferences will be held with the members from WBAMC and TTU. Fellows will be evaluated overall in terms of professionalism and ethical standards and specifically in terms of assessment, intervention, and follow-up of their projects. Typically, written reports will be made and new contracts negotiated as required.

Additionally, an Advisory Board will also be formed which will include local experts in CMH. Anticipated members will be those practicing, teaching, and consulting in the social sciences in the greater El Paso area. It is hoped that these members will not only advise us but also volunteer their talents to directly train the Fellows.

Methodology

In general the method of training will proceed from an initial period of theory and didactic exposure to supervised "hands'on" experiences to the eventual individualized practical application.

Initially, the Fellow will be exposed maximally to theory through coursework, reading, meetings with experts, and personal writings. The intent here is to insure maximum exposure to an interdisciplinary approach to problem solving and to master specific content. This represents the heart of the didactic phase.

Secondly, the Fellow proceeds into a supervised experience of direct consultation efforts. A behavioral program is agreed upon in accordance with the Fellowship Committee that meets the Fellow's individual needs and those of the organization. It is expected that the Fellow will work directly with local Mental Health Officials (to include non-military professionals) in a "hands-on" supervised approach. He will participate in on-going consultation projects.

Finally, the training will move toward a less intensively supervised experience. Here the transition from theory to application is completed. The Fellow works more on his own and brings to fruition a project he has developed which not only meets academic requirements, but has utility to the military community as well.

Qualifications for and Selection of Fellows

1. An AMEDD Commissioned Officer with a recognized commitment to the military (at least on INDEFINITE status).

2. A Psychologist with a minimum of one year's post-graduate, post-internship military experience (two years preferred) or a residency trained Psychiatrist with a minimum of one year's post-residency military experience (two years preferred) or a Social Work Officer with a minimum of one (preferably two) year-s post-Master of Social Work degree military experience.

3. Demonstrated leadership and management skills as assessed through personal interview and records review.

4. Professional excellence in individual specialty as assessed by academic preparation, OER's, and letters of reference/recommendation.

5. A verbalized committment to the principles of the community model as outlined herein.

6. The Fellow would agree to a two year "payback" commitment of service.

Summary

This paper has described a post-graduate Fellowship program in CMH. Experienced clinicians (psychologists, psychiatrists, and social workers) will be selected on rigorous criteria and specially trained in community skills. Objectives will seek to provide skills which will facilitate the servicing of broader populations and utilize primary prevention methods and techniques. The mothodology will include didactics, supervised experiences and individualized projects having military utility. This program will train a new military behavioral science specialty and provide a badly needed incentive for retention.

Bibliography

1. Goodstein, L., and Sandler, I. Using psychology to promote human welfare: A conceptual analysis of the role of community psychology. American Psychologist, 1978, 33, 882-892.

2. Luchsinger, V., Anderson, R., Hartzell, R., Rath, F., and Bleker, E. Community Psychology as a Viable DoD Activity. Proceedings of the Sixth Symposium, Psychology in the Department of Defense. U.S. Air Force Academy, Colorado, Springs, Colorado, 1978.

Community Mental Health Programs and Youth

Robert P. Anderson, Ph.D.
Lt. Andrea Delligatti, USAR
Texas Tech University
Lubbock, Texas 79409

Abstract

An integral part of the community mental health system consists of the young persons in the community. In a total systems perspective, these young people largely may be civilians, but the age range includes adolescents extending into the age ranges of the younger enlisted personnel. This discussion deals with the problems of the young people of the military community mental health system, and how they relate to the larger military community.

Any community mental health system serving military personnel must incorporate services for the family members of these employees. Dependent children of service personnel provide a significant segment of the population served by mental health workers in the Army. Poorly adjusted parents or dysfunctional marital groups tend to yield disturbed children. Similarly, exceptional children, such as those who are hyperkinetic or neurologically impaired can create stress on their parents. Independent of the etiology, disturbed children tend to increase stress on their parents as individuals and on the marital relationship. Thus, whatever the starting point or whoever is the "target" person, disturbed children effect the complex social-environmental system within which the service person is interacting. In this interacting system, a circular effect is created which has potential negative consequences on the service person's occupational effectiveness. It is our contention that the occupational efficiency of the soldier is markedly influenced by the mental health status of his/her dependent children. Therefore what happens to "the kids" is a legitimate focus of interest in community mental health.

Despite the demand for children's services in our "child-oriented" society, many mental health workers are trained primarily in individual diagnostic and treatment approaches of adults. These professionals are not really familiar with the specific needs of children; the social agencies utilized by children, such as schools or preschools; or the community resources available to children. The popular musical comedy, Music Man has a line sung by the hero in his admonition to a not so successful salesman, "...you've got to know the territory." This line has implications for many mental health workers as they attempt to work with disturbed children and dysfunctional families, "...you've got to know the territory." This is especially important since the disturbed child is usually living in an environment that contributes to his emotional disorder. In order to deal with and manipulate this environment effectively, we suggest that treatment programs for children have a broader community base.

In our opinion there is a deficit of knowlege regarding a) some of

PRECEDING PAGE BLANK-NOT FILMED

the specific problems of children, b) how these problems can be treated, and c) the system of social agencies with which the children interacts. We will outline very briefly some of the problem areas which are most relevant in considering the overall adjustment of "kids."

The number of married service personnel is increasing particularly among the enlisted ranks. Also the number of couples where both persons are in the service has dramatically increased. A natural consequence is more children among young couples who are not particularly prepared for the problems inherent in parenting. Young faimly units need parenting training as a first step in forestalling potential problems such as emotional deprivation neglect or child abuse. Developing an awareness of constructive alternatives to the criticize-hollar-spank approach is a necessity. The mental health worker with a community awareness can be the catalyst for the development of parental "in service" training. A special problem arises when inflationary pressures create the need for both parents to work. In this case and in that of the single parent, child care facilities become a critical necessity. In many settings, particularly in overseas areas or bases in rural settings, child care facilities are nonexistent or marginal at best. These marginal facilities are a breeding ground for psychological stress and the development of maladaptive behaviors. A unique problem relates to the married couple, both of whom are in the units, who have children. In case of an emergency or a normal necessity when both parents are "called," do they have adequate resources for the children? The mental health worker should be in a position of being the catalytic agent for the development of adequate child care facilities. Development can not proceed without evaluation of these facilities in terms of the effects of their programs on child development. During the preschool years, the child's world is dominated by interactions within the family. It is the regularity or irregularity of the family-dominated environment which will affect subsequent social adjustment. The mental health worker can be instrumental in providing a stabilizing influences in the family environment.

Upon entering school, the influence of the family environment begins to recede into the background, as the academic situation becomes the focus of the child's life. The effect of school depends on the child's personality and his ability to assimilate new experiences and modify previous perceptions. The majority of children can probably adapt at least superficially to a wide variety of school environments. Nevertheless, children with special problems, such as the mentally retarded, learning disabled and emotionally disturbed, cannot easily adapt to new situations. In order to deal with this problem, the mental health worker needs skills on two levels. The first is a knowledge of how the school systems operate. Public Law 94-142 has had a major impact on school programming for the exceptional child. Unfortunately there are both between and within state variations on how these programs are implemented. Expertise in consultation skills, geared to the realities of the particular school situation, are a necessity. The second skill relates to individual diagnosis. The professional must at least know how the diagnostic evaluation must be carried out and what to look for in a good evaluation.

Providing at-home treatment facilities for emotionally disturbed

children is also a critical need. Because of availability of the Champus resource, an inordinate number of service children are referred for services away from their home communities. We contend this is a result of the inability of the Army resources to deal with these children and to work with parents toward solutions allowing the children to function at home. In essence, it is easier for the family, the school, and the mental health worker to move the child off than to work toward providing care facilities within the community. There is strong evidence that children who can function in the "mainstream" or in education jargon, in the least restricted environment, make better long term adjustments than children identified as "problems" who are sent off for treatment.

The focus of problems change as children move toward adolescence. Again there is a need for resources which can help parents cope with a) normal adolescent acting out, b) emerging sexuality, c) problems of leisure, and d) drug use. Many smaller school systems have only two alternatives for teenagers, academic programs or sports. This lack of diversity enhances the teenager's need to engage in risk-taking behaviors in order to comply with peer pressures. So a result, the conflict between adolescents and their parents is escalated. The mental health worker can interrupt this dysfunctional pattern in several ways. First the mental health worker should act as a catalyst in setting up social activities for adolescents through various community organizations. Second, community programs should be developed in order to disseminate information concerning drug- and sex-related issues to both parent and teenagers. Finally, the mental health worker should catalyze group discussions among parents and/or teenagers as a means of decreasing the distance between adolescents and their parents and as a way of helping adolescents work through their confused feelings.

In conclusion there is a definite need for mental health services for children within the Army. The goal of a community-based mental health system would be to improvise its collaborative services with other agencies in the community to ensure maximum utilization of resources.

FAMILY AND MISSION CONFLICTS: THEIR IMPACTS ON READINESS

Mrs. Cecile S. Landrum
HQ USAF, ACS/Studies and Analyses
Washington, DC 20330

ABSTRACT

The changing military family has brought to the operations world a new dimension of conflict--that of the conflict of time and resources between the needs of the family and the needs of the mission.

INTRODUCTION

When one hires a person, one gets the entire person. In addressing the hiring of women in today's expanded and changing job market, employers and policy makers have focused on whether the woman can do the job--the 9-5 wrench turning. What they haven't focused on is the job holder as part of an organization that can make other legitimate demands on the employee and the subsequent and recent phenomena of the demands the employee and his/her family makes on the organization. Demands as they relate to the military mission require that military members always be ready to meet the mission. Therein lies the conflict.

WOMEN AS CAREERISTS

Forty years ago, women teachers could not drink or smoke in public and had to remain single. In the early 1960s, stewardesses also had to remain single--try that on today! Today, women are in all the professions, but they also marry and become mothers. Nancy Friday recognized the ensuing problem in her book, My Mother, My Self.

> Some women can combine full-time careers with being full-time mothers, but they are the superhumans among us, and you cannot base a rational society on all women being superpeople.
>
> She quotes Professor Jean McFarland: Some of our best women are choosing not to become mothers, not because they don't want to, but because they recognize they can't do both jobs well. (Friday, 1977)

There are clearly conflicts for women in their ability to manage a career, marriage, and/or family. Recent trends showed that women prefer employment and a career to full-time family and parental responsibilities. (Hayghe) When the role of women in the military was rapidly expanding in the early 70s, 43% of women adults were working outside their homes. (Murphy, 1978) The difference with the military is the mission requirement beyond the 9-5 job.

WOMEN WITH MILITARY CAREERS

While career motivations have been the basis for many women entering the professions, including the offier corps in the military, a large proportion of the young women are entering the enlisted ranks for reasons of security and job training.

Many reflect the very traditional values of wanting to marry and have families. They view their service time as just "time" until they do meet their mates and start raising families. Very few 18 year olds, men or

PRECEDING PAGE BLANK-NOT FILMED

women, think beyond these goals and in terms of long range careers. Major Pat Murphy refers to this phenomenon in her article, "What's a Nice Girl Like You Doing in a Place Like This"? by saying that "our society is still geared to a woman working only temporarily--until she gets married and has a family." (Murphy, 1978)

The military woman, unlike the civilian, however, has made a commitment to serve for a specified period of time. Therefore, many of her changing personal goals are temporarily subordinate to her commitment to the military commission; a major difference between military and civilian employment, even when the job skills appear to be the same.

Another group of women are entering the service knowing that their dependents will be the beneficiaries of health care, basic subsistence allowances and commissary and BX/PX privileges. Prior to 1969, women in the military could not have dependent children, so the combination of the 1969 policy changes and the 1973 lawsuit (411 USC 677: 1973) opened the door to women receiving benefits.

These changes established the environment for an entirely new population of military people for the personnel system to deal with. However, it appears that policy changes regarding growing numbers of women, job opportunities, assignment locations, marriages, families, and entitlements have all occurred at such a fast pace that compounding impacts have not been thought through.

Marriage and the Military Woman

For the most part, the young enlisted women, no matter what their reasons for joining the service, are of the age and inclination to be married. Their environment is conducive for making a match to another service member since they are so outnumbered by men, and are often away from their own friends and family, or at remote sites, or serving in foreign countries. In fact, the military, as an institution, has shifted from a predominantly single force during World War II to a married force. (Moskos, 1977)

Recent changes in military policies have also made it attractive for military members to marry. When both spouses are in the military, certain benefits become very appealing. For instance, both draw basic allowances. All the services are making great efforts to assign military spouses to the same bases, or at bases within commuting distances. This has resulted in some "convenience" marriages where a person agrees to team up with someone in order for one to get to the base of their desire.

The News and Courier of Charleston, SC, cites another example in which an E-3 stated that "he had never lived with his wife, also an E-3, or consummated the marriage." They agreed to marry so they could move out of the dormitories on base and collect their separate allowances for housing and food. ("A Business Agreement," 1978) While this could present some morale problems, identifying the basis for this is almost impossible and in the vast minority. Therefore, removing benefits would unfairly take away entitlements from the majority.

Regardless of why the women are marrying, the services are experiencing increasing numbers of inter- and intra-service marriages. Along with the new logistical, management, and human problems, there is the problem of impaired combat readiness. Non-availability caused by conflicts with

family responsibilities, or lack of job concentration by a spouse concerned with the safety, or whereabouts, of the other spouse in the unit, are especially serious, particularly in a contingency.

Pregnancy and the Military Woman

Approximately seven percent of the women in the Armed Forces are pregnant at any one time. However, due to the military personnel system, many jobs are constantly filled by a concentration of young women of childbearing age. As the numbers of women increase, the problems associated with pregnancies being concentrated in specific jobs could be difficult to deal with.

A young woman, particularly in the non-traditional fields, contemplating pregnancy and parenthood, must carefully examine her ability to assume the demands of military life. Her job requires full-time commitments, whether she is attending one of the academies, or is serving in a job which requires more than normal field duty, or time away from home. But many women want to be careerists, and therefore their pregnancies will add further responsibilities to an already demanding lifestyle. (Williams)

Although a survey conducted by Colonel John Williams, of the US Air Force Academy, indicated that the majority of the military couples were adamant about not having children, most of the couples interviewed were in their 20s. Many of the women I have spoken to, find themselves equivocating as they reach their 30s, when decisions on childbearing become irreversible. The new trend among career women, of having their first child in their mid-to-late 30s, demonstrates how fluid the family issues are.

THE IMPACT OF THE FAMILY ON THE MILITARY

Today's families include women with children, who enter career fields which require mission availability at all times; male single heads of households; joint spouse members with children; and military members' civilian spouses who seek careers of their own. Hamilton McCubbin, a sociologist at the University of Minnesota, states "It (the family) is expected to accept willingly the stresses of military life and maintain its traditional supportive but subordinate role as stress alleviator." (McCubbin et al, 1978)

However, societal changes, such as the emerging roles of civilian spouses with careers, are causing military families to openly question the organizational demands on their lives (i.e., frequent moves and separations). (McCullah, 1978) Admiral Elmo Zumwalt noted in a 1977 speech to a military conference on the family that the Navy is now finding it harder to compete with the family and that the wives are less willing to make the sacrifices they made in earlier years.

McCubbin, recognizing the demands the organization had made on the family in the past, asks the following questions: "How will these changing situations affect the military job assignments, family relocations, and extended separations? And will members of military families become less dependent on the system, more assertive of their personal and family needs, and less willing to subordinate their lives to the orders of the military establishment"? (McCubbin and Marsden, 1976)

Child Care and the Military

Comprehensive child care to meet military requirements involves different types of facilities and personnel. Child care plans during emergencies, extended periods, such as TDYs and exercises, and national emergencies need to be made before these situations occur. But many parents appear unaware of the need to really make different types of child care arrangements. Many believe there will never be a war, while others state that if the war comes, they are going home with their children. I have seen cases of one spouse, or the sole parent, not coming to work during an alert, or parents bringing their children to work during the alert, probably the worst time to have young children underfoot.

As dependent children get older (as women stay in service longer and enter with dependent children) their requirements for after school and during vacations are more demanding than those of small children. Sports and other creative programs are necessities in order to avoid problems resulting from unsupervised time. Many of these services, which have been a part of base life have been drastically cut back during periods of budget restraints. Often these services are viewed as luxuries, not necessities.

Thousands of single fathers, in far greater numbers than single mothers, (a reflection of the large percentage of males in the military) are being faced with child care concerns as women give up the nurturing responsibilities. These individuals include those widowed, single, divorced, under interlocutary decree, or whose marriages have been annulled. The children may be that of the parent, or may be adopted.

It is the opinion of some commanders (Tice) that the single fathers tend to be older than their female counterparts and more often have resolved their child care needs through a relative or help living in his home. Still, the situation exists.

There were some 3,538 male single heads of household and 1,107 female single heads of households in the Air Force in 1978. By September 1979, the number rose to 5,732 and 7,500 Air Force couples with dependent children.

IMPACT OF THE FAMILY ON READINESS

Interestingly, it was not too long ago that pregnant women, or mothers of minor children serving in the military, were considered as detriments to national security. In Struck vs Secretary of Defense, a legitimate government claim of national defense was established for automatic discharge on these grounds. (Struck vs Secretary of Defense, 1972) Although this rule has been overturned, in December of 1978 the Army announced that starting in January 1979 single parents could not enlist--they could be commissioned--and that service couples with dependents and single parents must demonstrate that their family responsibilities will not interfere with their military duties (except in some family emergencies). If this is not demonstrated, military members may be denied reenlistment, or can be involuntarily separated.

The Army has also said that they will not show favoritism to these service people by offering them headquarters jobs or by placing them in non-deployable units to the detriment of other soldiers. (Pentagram News)

CONCLUSION

What do family and mission conflicts mean to the Services. From the outset, the Services must be able to set the guidelines in terms of the conflicts of mission responsibilities versus family responsibilities. At the same time, all men and women entering the services must view their career potentials realistically, in light of extensive TDYs, immediate availability 24 hours a day, and all-out conflicts. They must consider how they will divide their time between career and family, and make child care plans accordingly. Additionally, policy changes must come that will alleviate the concerns and stress placed on the military member by the family.

"A Business Agreement," The News and Courier. Charleston, SC: 11 Dec 78, p7

"Dependent Care Plan Needed for Sole Parent, Couples," The Pentagram News. 30 Nov 78, p. 1-2.

Friday, Nancy. My Mother, My Self. NY: Dell Publishing Co., 1977, p. 257.

Hayghe, H., "Marital and Family Characteristics of the Labor Force," Bureau of Labor Statistics News. Washington DC: Department of Labor, Jul 78.

McCubbin, Hamilton I., and Marsden, Martha A., "The Military Family and the Changing Military Profession," paper presented at Inter University Seminar on Armed Forces and Society, Maxwell AFB, AL, October 1976.

McCubbin, Hamilton; Marsden, Martha; Durning, Kathleen; and Hunter, Edna. "Family Policy in Armed Forces," Air University Review. Vol 29, No. 6 (Sep-Oct 78), p. 46.

McCullah, R., from paper presented at 86th Annual American Psychology Association Convention, Toronto, Canada, September 1978.

Moskos, Charles. "The All Volunteer Military Calling: Profession or Occupation"? Parameters. Vol VI, No. 2 (1977).

Murphy, Patricia. "What's a Nice Girl Like You Doing in a Place Like This"? Air University Review. Vol 29, No. 6 (Sep-Oct 78), p. 74.

Tice, R. Dean, Major General (Cdr, 3rd Infantry Div) Remarks made as a panel member of Utilization of Women: Impact on Combat Readiness, Jun 79, 43rd Military Operations Research Society Symposium, West Pt, NY.

Struck vs Secretary of Defense, 409 USC 1071, 1972.

411 US 677: 1973. In 1973 the Supreme Court ruled as urconstitutional the law giving family entitlements as housing allowances, commissary privileges and health benefits to married servicemen and not their female counterparts in Frontiero vs Ricardson.

Williams, John. "Alternative Family Styles in the Military," paper presented at the 43nd Military Operations Research Society Symposium, June 1979

Human Development Counseling and the Air Force: A Successful Marriage?

Richard L. Percy
Academic Director - Peabody/Vanderbilt in Europe
RAF Upper Heyford, England

Earl A. Glosser
Director - Student Counseling Center
University of Virginia
Former Faculty Member - Peabody/Vanderbilt in Europe

Abstract

Two important trends, developing separately but simultaneously, have provided the impetus for the merging of a university-based graduate program in Human Development Counseling with a portion of the military community included in USAFE. These movements reflect the growing tendency within the Air Force to become more "people-oriented" and to place greater emphasis upon the effectiveness of its supervisors and managers, at all levels, and the professional trends in the behavioral sciences away from traditional counselor training programs and toward systematic human relations training models. This paper describes the program being conducted in England and in Berlin by George Peabody College of Vanderbilt University through a contractual arrangement with USAFE. The program rationale, curricular structure, and impact on the entire military community it serves are discussed. Preliminary evidence suggests many areas of mutual involvement and argues that the consortium has resulted in a happy marriage.

Historically, counseling and psychotherapy have been viewed as mystical and esoteric or at least highly specialized services of a clinical nature carried out on a one-to-one basis with an individual who has undergone years of intensive training. The purpose of this process is to cure or moderate mental illness. Over the past two decades considerable research findings (Carkhuff, 1969: Carkhuff & Berenson, 1977: Egan & Cowan, 1979: Gazda, Walters & Childers, 1975: and Truax & Carkhuff, 1967) have indicated that (1) the attitudinal qualities and skills generally associated with effective therapists are actually basic human relations/interpersonal relationship skills and that (2) it is possible to take lay persons or individuals with virtually no formal background and training in psychology and the behavioral sciences and train them in a relatively short period of time to significantly increase their levels of these facilitative traits and skills to a point where they can become effective helpers to those around them.

As a result there has been a very substantial movement in the helping professions toward a training model for practitioners which reflects the above findings. Such training models recognize the central value of communications and human relations training and view the skills learned through this training as generalizeable to any and all human relationships outside of the classical client-therapist relationship.

PRECEDING PAGE BLANK-NOT FILMED

It is now clear that there is nothing mystical about psychotherapy and that systematic human relations training programs offer not only the basic therapeutic skills necessary for those seeking to become professional helpers, but also offer an opportunity for trainees to develop skills which can lead to more personal effectiveness. The implications for such training models within a military community are staggering. Such training is capable of enhancing the effectiveness of military counselors, supervisors, managers, teachers, parents, and in fact everyone.

Background

In the fall of 1975, George Peabody College for Teachers, now known as George Peabody College of Vanderbilt University embarked upon a new bold venture as the result of a contractual agreement with the United States Air Forces in Europe (USAFE). Peabody/Vanderbilt initiated a full graduate training program in Human Development Counseling (HDC) in Europe leading to a Master's, Specialist, or Doctoral degree. The training model reflects the research findings discussed above and thus focuses on the training of human relations skills which will enhance the trainees effectiveness, not only as a skilled helper, but will provide the personal growth experiences necessary for becoming more effective human beings. Thus, through a sensitive balance of didactic and experiential learning, participants grow professionally and personally.

The Peabody/Vanderbilt graduate programs in HDC operate in seven locations in England and in West Berlin, mainly on U.S. Air Force installations. There are presently over 200 students enrolled in the program at the various degree levels. Some preliminary results and the implications for such training within a military community are discussed below.

Program

Using a contingent of six full time faculty members, all of whom have regular faculty appointments with Vanderbilt University and who are stationed in Europe for at least a year, the program offers some 72 formal courses per year at its various centers. In addition, numerous individualized independent studies are arranged to meet student needs, especially at the advanced graduate degree level. These experiences are supplemented by a formal weekend workshop series, which brings leading professionals from the states to Europe for intensive weekend training, and small doctoral seminars at Cambridge University.

The annual schedule consists of four 10-week terms with two or three different courses being offered at each center each term. Thus, participants can complete an entire degree program as quickly as 14 months and quite easily within two years. The skeletal structure for the program consists of six broadly defined curriculum areas - Foundations of Psychology: The Helping Relationship· Group Process: Life Styles and Career Development: Measurement and Appraisal of the Individual and Research Methodology. Students accumulate coursework in each of these six areas according to a prescribed pattern in pursuing degree programs.

With faculty moving from residence to residence up to three times per year, they are able to teach those courses which reflect their prime expertise without having to offer courses out of their area. It also enables most every trainee to come in contact with at least four and perhaps as many as seven different faculty members during their degree program. In an employment market where 300-400 applications are received annually for these faculty positions, there are few, if any, training programs stateside which expose their students to as many experienced, highly respected counselor educators. Such a feature argues strongly in favor of off-campus military education programs.

Results

Since its inception in 1975, nearly 500 individuals have participated in the Peabody/Vanderbilt overseas program in HDC. The number of graduates by degree levels is presented in Table 1 below.

Table 1

Overseas Graduates for Each Degree Program

Academic Year	Master's	Ed.S.	Ed.D.	Total
1975-76	31	2	0	33
1976-77	77	14	1	92
1977-78	74	13	2	89
1978-79	66	9	4	79
1979-80*	65	9	3	77
Total:	313	47	10	370

* Projected

The vast majority of these individuals are on active duty in the military (the greatest number are in the Air Force but there are a few members of the Army and Navy also participating). The non-military members are mainly dependents but also include DOD school teachers, counselors, and administrators and also a few other non affiliated individuals living in Europe.

While research studies involving this overseas training model are just now starting to evolve, one study which has been on going since 1975 has attempted to make some comparisons between on-campus students and overseas students in HDC at both the beginning and at the end of a prepracticum experience. The questions being pursued were whether overseas students entered the prepracticum with substantially different skill levels from those on the home campus in Nashville and if not

whether the off-campus training would result in differential skill development for the two groups. The measures used were (1) a Self-Concept Q-Sort, (2) Human Relations Response Inventory, and (3) an Index to determine the perception of facilitative and non-facilitative responses. The data collected thus far suggest that both groups are comparable on all measures both at the beginning and at the end of prepracticum training. The implications is that even though the students in the overseas program are in a military setting, tend to be older, work full time, and receive all of their training off campus without the benefit of modern training laboratories, there appears to be no significant difference in the quality of these two populations initially and that the training experience overseas is as effective as the one experienced on campus - with regard to these measures. The data related to these findings are presented in Table 2.

Table 2.

A Comparison of Mean Pre-and Post-Test Scores of Overseas and On-Campus HDC Students on Prepracticum Instruments

Instrument	Nashville				Overseas			
	Pre		Post		Pre		Post	
	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$	N	$\bar{X}$
Self Concept Q-Sort	34	90.8	32	75.1	81	100.8	48	77.4
Helping Relations Inventory*	35	48.8	23	37.9	78	45.9	56	36.6
Index of Perceptions	19	.59	12	.48	86	.57	39	.47

*Score for understanding response only

Peabody/Vanderbilt overseas graduates have found their skills and expertise valuable in a number of ways including teaching in the public schools and at institutions of higher learning such as the University of Maryland, Park College, St. Martin's College, Pacific Lutheran, City Colleges of Chicago, and Boston University. They are authors of books, they conduct workshops and training courses on such topics as Women's Issues, Marriage Enrichment, Rational Approaches to Weight Control, and Psychopharmacology. They have received appointments to clinical psychology internships at both Fitzsimmons Army Hospital and Wilford Hall at Lackland AFB. They have been named outstanding members of their wing, the Third Air Force, and USAFE. The training in HDC has enabled physician assistants to incorporate effective counseling techniques into their medical practice and has also enabled the chief anasthesiologist at the Army hospital in Berlin to develop an effective pre and post operative counseling program for his patients. HDC graduates have received faculty appointments to the U.S. Air Force Academy, are employed as licensed psychometrists in a community mental health center,

have established child care centers, are engaged in pharmeceutical research, and serve as a Dean of Students. A number of enlisted personnel have secured appointments to OTS, partly based on their educational achievements and an undeterminable number of military officers and non-commisioned officers have received promotions thanks at least in part to their graduate degree.

Discussion

On a limited but important scale, the Peabody/Vanderbilt overseas training program in HDC helps to provide a vital support system for the military community. Some of the significant areas in which its impacts are discussed below.

There are many roles and duties within the military which relate directly to some facet of the helping professions. Mental Health clinics Social Actions, CDAAC, and the chaplaincy are the more obvious ones. All of these areas of military life are well represented in the Peabody/Vanderbilt student body as these individuals look for opportunities to supplement their previous training and thus to enhance their effectiveness as a skilled helper.

The completion of the field-based practicum experience represents the culmination of professional training for those aspiring to be counselors. During each year somewhere between 50-85 individuals are involved in practicum training under the supervision of a faculty member. The practicum focuses on the daily coping needs of normally developing individuals and a large percentage of individuals served through this practicum experience are members of the military community.

As today's Air Force strives to become "people-oriented", it relies heavily upon the effectiveness of its supervisors and managers, at all levels. The most recent development in the overseas program, has been the inclusion of a new option - Track II - which provides specialized learning experiences in supervision and management, largely on a weekend format, for those individuals who would like to apply their human relations/communications skills directly within their duty settings.

These represent areas in which the Peabody/Vanderbilt HDC program has made important contributions to the individual and collective needs of the members of USAFE. While the time is right for collecting empirical data to substantiate these contributions, there already exists considerable evidence that it has been and is a successful marriage for both parties.

References

Carkhuff, R. R. Helping & Human Relations. Vol 1: Selection and Training. New York: Holt, Rinehart & Winston, 1969.

Carkhuff, R. R. & Berenson, E. G. Beyond Counseling & Psychotherapy. New York: Holt, Rinehard & Winston, 1977.

Egan, G. & Cowan, M. People in Systems: A Model for Development in Human-Service Professions & Education. Monterey Ca: Brooks/Cole.

Gazda, G. M., Walters, R. P. & Childers, W. C. Human Relations Development: A Manual for Health Sciences. Boston: Allyn & Bacon,Inc.,1975.

Truax, C. B. & Carkhuff, R. R. Toward Effective Counseling & Psychotherapy: Training & Practice. Chicago: Aldine Pub. Co., 1967.

The Effects of Mobility on the Military Work Force and Their Families

Patricia Cooney Nida, Ph.D.
Director of Nida Training and Development Group
Escondido, California 92027

Abstract

Mobility has become a way of life for military families. This paper presents a critique of the research in the field and also enumerates a number of issues which cause grief and stress for some mobile families.

The expression, "Moving is a fact of life for many American families," has become a cliche in the literature on mobility and life-change stress. Not only in the military but in the corporate work force as well, families have been asked to pack up and go where the organization needed them. Although historically American families have always been fairly mobile, only since World War II have serial nuclear family moves become common. Traditionally these highly mobile families were expected to pack without a murmur of complaint and "adjust" to the new environment. The cost, in terms of individual suffering, was not addressed.

Several authors, notably, Toffler in Future Shock, Packard in Nation of Strangers, and Seidenberg in Corporate Wives-Corporate Casualties?, chronicled the effect of constant moving upon American families and society. Their writings clearly describe a nation where one in five families move each year, and the resulting lack of personal involvements or community roots on the part of the moving families.

During this period of high corporate and military mobility, a concept developed describing chronic depression and general isolation among many women in America which had a foundation in psychological theories describing women as simply prone to endogenous depression or filled with envy at the male role. Women who were not able to "cope" with constant moving or who wished to develop something in their lives that was broader than their role as a wife and hostess were considered by many to be showing symptoms of at the least, maladjustment and at the worst, mental illness (Chesler, 1972). Concurrent attitudes related to the physical health of women. Doctors reported a large number of women coming in with vague complaints which seemed to have no organic cause. Many physicians seemed to see these complaints as simply sex-related maladies or an attempt by lonely women to get attention.

Toward the end of the 1960's and early 1970's, a change in values and subsequent behavior started to occur in the population of the United States. The willingness to subjugate ones self to the goals of an organization, to sacrifice happiness and stability in life in order to get ahead in a career, to devote life energy to work rather than personal or family pleasures, all came into question. "Self-actualization" replaced "adjustment" as a catch phrase to to describe a healthy personality.

This new emphasis on self-actualization, however, is producing a great deal of role conflict, guilt, and anxiety for some women. Brought up to think of a "good wife" as one who finds total joy and fulfillment in serving her family and following her husband's leads, these women now find themselves dissatis-

PRECEDING PAGE BLANK-NOT FILMED

fied with the very life they have been programmed to cherish. They are frightened by the tension they feel and blame themselves, not their environment, for their problem. Many of the women are now facing these very painful personal issues and one life-change event that exemplifies this conflict is geographic relocation.

The literature on the effects of mobility on families is incomplete and inconclusive. Although there is agreement in clinical observations that a high rate of mobility does contribute to a sense of rootlessness in our society (Toffler, 1971; Packard, 1974) and to a disorientation and depression among the highly mobile wives (Seidenberg, 1975; Weissman, 1972) comprehensive experimental or even correlational data is lacking. In research that does exist there is considerable contradiction with some studies stating that no relationship can be found between moving and psychological stress and other studies indicating that there seems to be a direct relationship between the two.

Much of the research has methodological problems. For example, many studies use a small n and some of the data is old. Attitudes on wive's willingness to subjugate herself to the needs of her husband may be out of date if they were collected prior to the advent of the womens' movements. Economic factors of the last few years have changed so drastically it is impossible to generalize with certainty from studies conducted prior to the recent spiraling cost of living. Also, studies done in the corporate sector reporting attitudes of employees whose companies underwrite the economic loss of buying a new house or the cost of a spouse's employment agency fee cannot be presumed to have external validity for the military population where these services do not exist. In some studies the text reports the findings in a way that while statistically accurate, misses the important practical information offered by minority views. For example, while the text may report "most of the wives were happy with their moving experience," the data indicates that 30% of the women reported they were not happy with their moving experiences. Several studies of family adjustment use the reports of the husband stating what he considers to be the attitude and mental adjustment of his family to moving. His perception of the family's adjustment may be biased by his own needs to justify his action in constantly uprooting his family. Finally, a number of authors conclude that although a family may be experiencing a substantial amount of moving related stress, it is common for them to report symptoms only and not to recognize them as the result of a move. Therefore, studies which ask direct questions as to whether moving has been the cause of family problems may be invalid.

The work of Thomas H. Holmes, M.D., and Richard H. Rahe, M.D., indicates that any life-change produces stress that, in sufficient quantities, can have a direct impact upon physical and psychological health. Using their Life-Change Scale, events which are present in relocation are of sufficient magnitude to put an individual at risk for stress related illness. Further research indicates that social support offered by friends, family, and neighbors can be effective in reducing the risk of stress related illness (Kaplan, Cassel & Gore, 1977). This is the very support which is often stripped from a relocating family.

Dr. Robert Seidenberg in his classic book, Corporate Wives-Corporate Casualties?, describes the effects of the high rate of mobility upon transient families. It is Seidenberg's premise that the primary breadwinner (ordinarily

the husband) withstands numerous transfers with minimum sense of loss because relocation for him means a transfer into a work position of equal or higher status than the one he left. However, Seidenberg feels that the wives and children in families where repeated moves are the norm, pay an enormous price for their constant relocation. Their credentials are not tied to a job that is changing with them, but are more likely to be tied to activities which are left behind in their old community. Friends left behind are not easily replaced. Seidenberg, in his psychiatric practice, reports treating numerous women who had grown depressed through fatigue at the challenges involved in reestablishing themselves in each new community.

Although most of the literature does concentrate on the wife and children, there is, in the literature of management, increasing concern over employees' stress (Levinson, 1975). The theories of conjoint family therapy clearly indicate that when one member of a family is stressed there is a profound effect upon all other members. At the minimum, we could presume that an employee whose family was feeling disorientation and disassociation due to moving would feel pressure from their pain. In addition, many moves due to promotion, or simply assignment to a new job, call upon the employee to "break-in" to a new work environment. We are now starting to see literature on mid-life crisis and its effect upon mid-career employees (Levinson, 1978). Mid-life crises often seem to be connected to a significant life-change event and for some employees a move to a new assignment may provide this event.

There is some indication in the literature that the successful rerooting of a family in a new community depends largely upon the "positive attitudes of the wife." Herman and Werbel (1978) report that children who have difficulty in relocation are likely to have mothers with a negative attitude toward moving. However, they point out correlation does not imply causation. It is possible that mothers whose children have difficulty moving may develop a bad attitude toward the move as a result of the problems encountered by their children.

Related to this is the notion sometimes heard that strong families or marriages that are "good marriages" can withstand the rigors of moving while marriages that are already in trouble or unsound will be ended by the stress of relocation. This argument presumes that if love is strong enough, it will be invulnerable to stress in the environment. Many married couples are no longer willing to "own" the guilt that comes with this theory. They state that their marriage is just fine, it is the strain of moving that is causing their problem and in order to regain the stability that they feel is possible in their marriage they no longer are willing to move.

Almost one-half of the women in this country work today. In many cases, they have professional careers which require several years of college training, but in other cases, women even without formal education who have established seniority in businesses, or have developed skills unique to their community, are finding that they too have a career orientation. Many are finding it extremely painful to give up their hard earned positions in order to follow their mate.

Many military families now find it takes two full salaries to maintain their households and any suggestion that they lose one salary so that the other member can transfer produces anxiety for both parties. Many women are aware that American society assigns status in accordance with economic income. These women are often unwilling to give up their new found "power" by

sacrificing their jobs and again being "dependent" upon their husbands.

Economics and double digit inflation are also having a direct impact upon families who are apt to move. These families have been taught to value material goods and are unwilling to take a drastic change in lifestyle, but are unable to stretch their income to meet a new housing market. Anxiety over this loss of economic security and this declining lifestyle may well produce symptoms of tension in military families.

People raised in different areas of the country or in different ethnic groups may have developed deep, if unconscious, values as to what type of home environment is "right" or signals success and "the good life." In order for two married people to build a home that they consider mutually satisfactory, they must often go through a period of negotiations based on each person's values brought forward from childhood. In many cases, the stress of finding just the right home reflects a great deal more than just practical decisions reached through the adult ego state. Rather, such conflict is the result of "parent tapes" and ties to feelings of self-worth.

Also, I believe aesthetic values are involved in a person's ability to reroot in a new climate. Although a person may understand consciously that Mt. Ranier is a beautiful sight, looking through the rain clouds of Fort Lewis in order to glimpse that mountain may do nothing to nourish the heart of a person from the southwest. In other words, the hyacinth for one's soul may well be defined by the aesthetic values ingrained in a person at a young age. Any move away from that home environment may be a move from the nourishment of the soul. This loss of nourishment may be manifested by a deep longing as expressed in the following poem written by Sandra Buehrer, a woman from Colorado, while living on an Army Post in the Panama Canal Zone.

I find myself in a land that is and will remain so alien to me.
Its beauties are not mine.
The things that I have cherished from my childhood belong to me:
The snowfall in a twilight
The cactus blooming radiantly in Spring
A rabbit journeying across my yard
And wheat ripening as far as eye can see.
The jungle greenness dripping wet recalls the autumn rains and winds.
The picnics by the sea bring back the campfires in the canyons and the echoes there.
I watch the pelican who fishes so adeptly by the shore and remember meadowlarks that called
And eagles soaring in their observation of my land,
My home,
The place where I belong.

So what are the answers to these complex issues which are impacting the lives of today's highly mobile military force? Certainly the solutions will be complex and probably will differ for each family. However, the awareness that there are indeed many legitimate reasons for families involved in repeated relocation to experience stress is at least the first step in the right direction. It opens the door to careful planning to reduce mobility rates and, more important, it enables helping professionals to hear clearly the grief that their clients may be feeling for a lost home.

References

Chesler, P. Women and Madness, New York: Doubleday, 1972.

Herman, J. & Werbel, D. The Effect of Job Transfer on Employees and Their Families, Washington, D.C.: Employee Relocation Council, 1978.

Holmes, T. & Rahe, R. "The Social Readjustment Rating Scale." J. Psychosom. Res., 1967, II: 213.

Kaplan, Cassel, Gore. "Social Support and Health." Medical Care, 1975, Vol. 15.

Levinson, D. The Seasons of a Man's Life, New York: Alfred A. Knopf, 1978.

Levinson, H. Executive Stress, New York: The New American Library, Inc., 1975.

Packard, V. A Nation of Strangers, New York: Pocket Books, 1974.

Seidenberg, R. Corporate Wives-Corporate Casualties?, New York: Anchor Books, 1975.

Toffler, A. Future Shock, New York: Bantam Books, 1971

Weissman, M. & Paykel, E. "Moving and Depression in Women". Transitions/ Society, 1972, 9 (9).

Thanks to: Michelle Egan, Research Assistant

Readability of Air Force Publications

Lydia Razran Hooke
Air Force Human Resources Laboratory
Lowry Air Force Base, Colorado 80230

Abstract

A recent regulation (AFR 5-1) requires AF writers to match the reading level of their publications to that of their intended audiences. The present study investigated whether rewritten regulations submitted as complying with AFR 5-1 were actually comprehensible to their intended readers. Reading levels of submitted regulations in seven career fields were rechecked using the required FORCAST formula and then appropriate CLOZE tests were developed and given to field members. In all, 1471 such tests were administered. None of the writers produced a regulation with a reading level much below that of 11th grade, while the mean level in three of the fields was below 10th creating cases of mismatch or literacy gap. In all three of these cases, CLOZE comprehension scores were below the established criterion level.

AFR 5-1, intended for writers of AF publications, prescribes and makes mandatory a procedure for determining whether a text is too difficult for its intended audience (i.e., whether there exists a negative literacy gap), as well as the measures to be taken if this is the case. Writers of AF texts are required to certify that the reading level matches that of the target audience. To make this certification, the reading level of a newly written or revised publication must first be determined by using the FORCAST formula, which is based on the proportion of one-syllable words in a text. If the grade level of the publication exceeds that of the intended audience, the writer must rewrite it until this is no longer the case.

The Technical Training Division of the Air Force Human Resources Laboratory undertook the study being discussed here in order to evaluate the effectiveness of AFR 5-1 as a standard for writing Air Force regulations. In particular, it was intended to determine whether regulations written in accordance with this regulation were comprehensible to their intended user audiences.

The central portion of this research began with the selection of regulations from seven different Air Force career fields, which had been rewritten in accordance with AFR 5-1 and submitted to AFHRL. The next step involved ascertaining whether these regulations would indeed be comprehensible to those who had to read them in performing duties. To answer this question, two procedures were used. First, selected publications were rechecked using the FORCAST formula to determine how successful writers had been in meeting their goals. Second, and more important, the comprehensibility of the publications to appropriate populations was measured. The instrument selected to effect this measurement was the CLOZE procedure. This procedure involves deleting every fifth word in a prose passage. Subjects are required to fill in the blanks left by these deletions. Their ability to supply the exact words deleted has been shown

PRECEDING PAGE BLANK-NOT FILMED

to be a good measure of how understandable (or readable) the passage is to them.

It was further decided to use a criterion-referenced approach to evaluate the meaning of the CLOZE scores. In other words, rather than compare new and old versions of the same regulation, an absolute standard of comprehension was specified, and performance above standard was considered acceptable. Here the 75% comprehension level was deemed to be an adequate criterion, and a CLOZE score of 40% (roughly corresponding to this criterion) was selected as the standard. The mean performance of subjects in each career field relative to this standard could then be assessed. This approach to interpretation of the data would provide direct answers to the question of whether the publications sampled met the standard for being sufficiently comprehensible to their intended readers.

Once the seven regulations were selected, it was necessary to devise measures of their comprehensibility to appropriate readers. It was decided to use the CLOZE procedure to develop such tests. This procedure entails deleting every nth word in a prose passage. Subjects are required to fill in the blanks left by these deletions. Their ability to supply the exact words deleted is taken as a measure of how understandable (or readable) the passage is to them.

In addition, seven vocabulary tests were constructed by asking the writers of the regulations selected for the study to provide a list of the career specific terms most likely to be well known to members of that field. These writers also supplied the correct definitions for each term. Each set of vocabulary items and definitions was then made into a matching test. Since the FORCAST estimate depends on the number of monosyllabic words in a passage, regulation writers had asked to be allowed to count well-known career specific words of more than one syllable as one for the purposes of the estimate. The rationale for this part of the experiment, then, was that if such words are indeed familiar enough to warrant being considered equivalent to monosyllables, personnel ought to demonstrate nearly perfect performance on vocabulary tests of such words in their own fields.

Approximately 900 Air Force personnel on 13 bases belonging to 15 major commands were tested in this study.

The first result of interest concerns the comparison of the publication writers' estimates of their regulations' RGL and the corresponding FORCAST computation performed by AFHRL staff members. For the regulations used, writers underestimated actual RGL of their publications by more than one grade level in four out of the seven cases.

After testing, the mean CLOZE scores for each of the seven career fields were calculated. These are displayed in Column 5 of Table 1. These scores represent the average across from two to eight tests developed for that career field.

Recall that the criterion set for adequate comprehension was a CLOZE score of 40% or better. It may be seen from Column 5 of Table 1 that the mean scores from only four of the seven fields (viz., Manpower Management, Personnel, Fuels, and Information) are above criterion. If the 95% confidence interval is considered, the criterion performance of subjects

in the Information field is called into question. Where there is no negative gap in a field (viz., in the Personnel, Fuels, and Information specialties), subjects have achieved above criterion scores on their CLOZE tests. Conversely, in three out of the four fields where there does exist a negative gap (i.e., Administration, Printing, and Nondestruction Inspection) performance is below 40% standard. The sole exception is the test performance of Manpower Management subjects.

Career Field	Reader RGL	Text RGL	Lit Gap	CLOZE Score	Vocab Score
Administration	9.1	11.6	-2.5	30.8%	70.2%
Printing	8.9	12.6	-3.7	27.5%	63.8%
Manpower Management	11.3	12.3	-1.0	50.4%	73.5%
Personnel	10.5	10.4	+0.1	44.1%	87.2%
Nondestructive Inspection	8.5	10.8	-2.3	32.0%	94.6%
Fuels (Officers)	14.0	13.2	+0.8	48.0%	98.4%
Information	12.7	11.5	+1.2	41.4%	63.3%

Table 1

Table 1 shows that for every career field where the members perform below criterion, the target RGL is below 9.6. Conversely, the target reading grade level for the fields where subjects do perform adequately is in every case above 10.4. Thus, the target reading level of the user audience is confounded with the variable of literacy gap in six out of seven cases. In other words, one cannot be sure that the failure to reach criterion was directly related to literacy gap or merely to the fact that the target RGL was low in an absolute sense. In the one case (Manpower Management) where target RGL and direction of the literacy gap yield conflicting predictions, the prediction based on target RGL is confirmed.

Data on vocabulary scores are given in Column 6 of Table 1. The hypothesis being tested was that these terms would be responded to by field members with virtually perfect accuracy. If a rather lenient standard of 90% is identified with this claim, it can be seen that only members of the Fuels and Nondestructive Inspection fields have attained this criterion. (It is noteworthy that in the case of Nondestructive Inspection, high vocabulary scores coexist with low CLOZE scores.) If a standard of 75% correct is set as the minimally adequate level of job term knowledge, only one more field, Personnel, shows acceptable performance. Four of the seven career fields failed to achieve even minimally

acceptable performance.

The first conclusion to be drawn from this study is that a substantial number of those responsible for rewriting AF regulations according to AFR 5-1 are not able to use the FORCAST formula to estimate accurately the RGL of their publications, nor were they hitting the target level of their intended audience. The data indicate that this was especially true when the target RGL was low, below tenth grade. Rewriting AF publications to have a low RGL is indeed an extremely difficult task. Most publications must contain a large number of multiword, polysyllable terms which are fixed, i.e. may not be "written down." After all, one may not paraphrase the title of a regulation or rename a specific piece of aircraft equipment. When target RGL is relatively high and little or no rewriting is required, this does not cause problems. However, none of the writers whose publications were used in this study were able to produce a document with an estimated RGL (as computed by AFHRL) below tenth grade in spite of the fact that they were required to do so to match the reading ability of the intended users.

The next set of conclusions to be drawn concern the CLOZE scores. First, the data indicate that when target RGL is relatively high and there is no negative literacy gap, comprehension of career field relevant regulations is acceptable, as indicated by the fact that CLOZE scores are above the 40% criterion. This would seem to be true even in the case of the Information field, where vocabulary scores reveal a striking lack of familiarity with job-specific terminology.

On the other hand, where reading level of the audience is low (< 9.5), there does tend to be a literacy gap and CLOZE scores are below criterion, indicating inadequate comprehension. This is true even in the case of the Nondestructive Inspection field where vocabulary scores indicate high familiarity with career specific terms.

As stated in the results section, in the one case (Manpower Management) where there did exist a negative literacy gap but user RGL is moderately high, the prediction following from the target level was confirmed, i.e. performance was above criterion. Thus, this study provides no positive evidence that literacy gap has an independent effect on comprehension, over and above the effect of low reader RGL. It can be noted here that in a related AFHRL study (Kniffin, et al) which used a factorial design, there was a small but significant effect of literacy gap on comprehension. The presence and magnitude of the effect of negative literacy gaps on AF publications would have important implications for policy regarding these publications. For this reason, it is recommended that a study be performed which separates the effects of literacy gap and low reader RGL, using materials and CLOZE tests comparable to the ones in the present investigation.

The main conclusion to be drawn from levels of performance on the vocabulary tests is that there is no basis for counting supposedly well-known career specific terms as one syllable words for the purposes of FORCAST estimates of the RGLs of AF regulations. The justification offered for so doing was that such terms are, to members of the appropriate fields, as familiar and easy to understand and deal with as common one-syllable words.

The results show that this is far from being generally true. As we have seen, AF publications directed at personnel with relatively low reading ability were generally too difficult for their user audiences. Allowing writers to count career specific terms as monosyllabic would simply further exacerbate this problem. Moreover, the Nondestructive Inspection field provides an example of personnel showing high familiarity with career terms, yet indicating by their CLOZE scores that they were not comprehending regulations written for them.

Finally, the low level of performance on vocabulary tests in four out of the seven fields demonstrates that the regulation writers, who submitted these terms as well-known, were overly optimistic with regard to the actual abilities and knowledge of their intended audiences.

As was the case with CLOZE scores, vocabulary scores increased significantly with skill level in the majority of the career fields investigated. This is some rough indication that the writers did at least choose field appropriate words. This finding also provides an encouraging confirmation that job appropriate knowledge does increase with skill level.

Implications for the Air Force

1. Writers of Air Force publications and regulations should get additional training in the use of the FORCAST formula and in ways to decrease the RGLs of their texts.
2. For the purposes of FORCAST RGL estimation, "familiar" multisyllabic words should not be counted as words of one syllable.
3. The policy of writing to target audience should be continued, but should not be enforced rigidly. Given the relatively crude ways employed at present to estimate literacy gap, it is probably not a good idea to insist that writers hit their targets with a great deal of precision. Additionally, the practical problems involved in simplifying materials below tenth grade level may be insurmountable.

Meeting Reading Requirements for Military Service

Clay E. George and Adriana D. Barrera

Texas Tech University

Abstract

A case is made that, in the short run, many military recruits will be unable to read sufficiently well to meet service needs. Original data are presented from a program designed to help language-disadvantaged young adults. Comparisons with data from earlier programs were found to produce gains in reading rate, but closer analysis revealed that little of the gain was made by the poorest readers. Such programs also produced negative correlations between rate and comprehension change scores. A program emphasizing comprehension development was successful in producing comprehension gains, without the negative correlation between rate and comprehension change scores.

Mathews, Valentine & Sellman (1978) have published data showing that qualified applicants to the services have a median reading grade level of 10.2. Although the simplest of service instructional materials are written at the sixth grade level, some 11% of the qualified applicants could not read at even that level. Given an overall shortage of applicants, these data indicate that there are severe limitations on the ability of the services to assign personnel to training schools. The fact that skill requirements are steadily increasing further complicates the problem. Banks, Berghage & Paje (Note 1), for example have reported a correlation between a pass-fail dichotomy and global reading grade level of $r(98)=.22$ in one service school for which students had already been highly selected.

If the services contract for the development of basic educational skills in applicants before enlistment, there are a number of points to be considered in stating contract terms and in evaluating outcomes (Williamson, Prost & George). If global reading grade level becomes a criterion of contract compliance, the all too human tendency may appear of achieving the promised result by producing "gee-whiz" gains in the easily manipulated rate of reading component without producing literary in the student. While it is true that reading rate, vocabulary and comprehension scores are all positively correlated with one another (and with grade level) "in nature," it does not follow that a training-induced change in rate will produce corresponding changes in vocabulary or comprehension. We have calculated a rank-difference correlation, $r(8)=-.58$ between mean rate and comprehension changes in class sections over a semester long training program in a military college and a Pearson correlation, $r(54)=-.35$ between changes in individual students over a similar time span in a private university. Descriptions of these programs are available elsewhere (Kingston & George, 1955; George, 1962).

PRECEDING PAGE BLANK-NOT FILMED

It also happens that mean rate changes resulting from a program may be comprised of very large gains scored by those who read fairly well and modest gains by those who read very poorly. The theoretical curves presented in Figure 1 are based on data collected over a period of 25 years. The upper curve is based on pre-post results from three programs (Ns of 73,55 and 20). The lower curve is based on two suitable control groups (Ns of 87 and 40). Actual rate scores in words per minute are not shown since these values depend upon the specific test used to measure them.

In general, however, the trends of the data indicate that programs which rely upon rate gains as evidence of effectiveness tend to produce a small advantage for extremely poor readers, no advantage for low average readers and massive gains for initially fair to adequate students.

Finally, commercially packaged reading programs tend to be based on the questionable assumption that changes in behavior resulting from training on special devices and materials will transfer to other reading situations. Tinker (1958) has shown that this is not so in the case of tachistoscopic training. George (1955) reported results indicating that training on reading paces will transfer to non-paced situations only under very carefully contrived conditions. In general, training produced changes in such peripheral events as saccadic eye movements or span of apprehension do not necessarily or automatically affect central processes involving the acquisition of information from textual materials.

From the above considerations, we hypothesized that reading comprehension improvement could be demonstrated in a program emphasizing vocabulary and comprehension development with secondary attention to rate. A secondary hypothesis was that rate and comprehension change scores would not be negatively correlated as was the case with the rate emphasis program.

Method

Subjects. Freshman students in one-half of the first semester English course sections during one summer session at Texas Tech University were offered the opportunity to participate in a reading program. Of those who were offered the training, 20 completed three or more hours and became the trained group. There were 40 students available from the remaining sections to serve as a control group. The first hypothesis was tested with data from 15 trained (3 to 20 hours) and 15 matched control subjects.

Procedure. Each training period began with a vocabulary development exercise. This portion of the training covered affixes as cues to meaning, dictionary use and contextual cues. The students then read a short article, timed their reading rate and answered comprehension questions. Next, there was a short group discussion of what the author had meant to convey and how he/she accomplished that end. Finally, the students were shown how to use a reading pacer, set just a little faster than their usual rates, and allowed to read a second article.

Results

Both groups tended to show gains in rate and vocabulary scores as noted in Table 1. The hypothesis that the trained group would gain in comprehension as a function of the program was supported in part. More convincing support would have required the experimental mean gain to have

Table 1
Mean pre & past test reading scores

Group	Test	Rate	D	Comp.	D	Vocab.	D
E	pre	213.13	+40.14	37.53	+3.80*	31.80	+2.87
	post	253.27		41.33		34.67	
MC	pre	221.80	+34.37	38.80	-0.40	32.60	+2.13
	post	256.17		38.40		34.73	
$\bar{E}$-$\bar{C}$			+5.56		+4.20		+0.74

*p .05

been significantly greater than the matched control mean gain, which was not the case.

Dynamic correlations for rate and comprehension change scores were calculated separately for the total trained and undertrained groups. For the trained group $r(18)=.22$, ns, while for the untrained $r(38)=.11$, ns. Although the correlations were not significant, they were not negative.

Discussion

It would be nice if one could select or train people for the ability to read at a given grade level and then be able to expect them to learn from printed material written at that level. Unfortunately, reading is a much more complex process than is often assumed to be the case. A useful degree of reading ability depends upon sight recognition of commonly used words, a capacity for learning new words in context, and the ability to follow the author's logic: all in a reasonable time span.

The training necessary to achieve functional literacy in adults with some prior reading ability must emphasize cognitive skills. Our data suggests that the goal of understanding be kept in the forefront of reading programs just as it is in most other training programs.

Reference Note

1. Banks, W. W., Berghage, T. E. & Paje, M.A. "A pilot evaluation of reading achievement test scores and other variables as potential predictors of successful student training in explosive ordinance disposal." Paper presented, U.S. Naval School, Explosive Ordinance Disposal, Indian Head, Maryland, September 1977.

References

George, C. E. "An evaluation of selected methods of using reading pagers." In O. S. Causey & A. J. Kingston (Eds.), Evaluating College Reading Programs. Fort Worth: Texas Christian University Press, 1955.

George, C. E. Emotional Adaptability and Intellectual Efficiency. Unpublished doctoral dissertation, University of Houston, 1962.

Kingston, A. J. & George, C. E. "The effectiveness of reading training at the college level." Journal of Educational Research, 1955, 48, 467-471.

Mathews, J. J., Valentine, L. D. & Sellman, W. S. "Prediction of reading grade levels of service applicants from Armed Services Vocational Aptitude Battery." Proceedings of the Military Testing Conference, 1978, 16, 494-506.

Tinker, M.A. "Recent studies of eye movements in reading." Psychological Bulletin, 1958, 55, 215-231.

Williamson, G. A., Prost, J. M. & George, C. E. "The professional psychologist as evaluator." Professional Psychology, 1978, 9, 308-314.

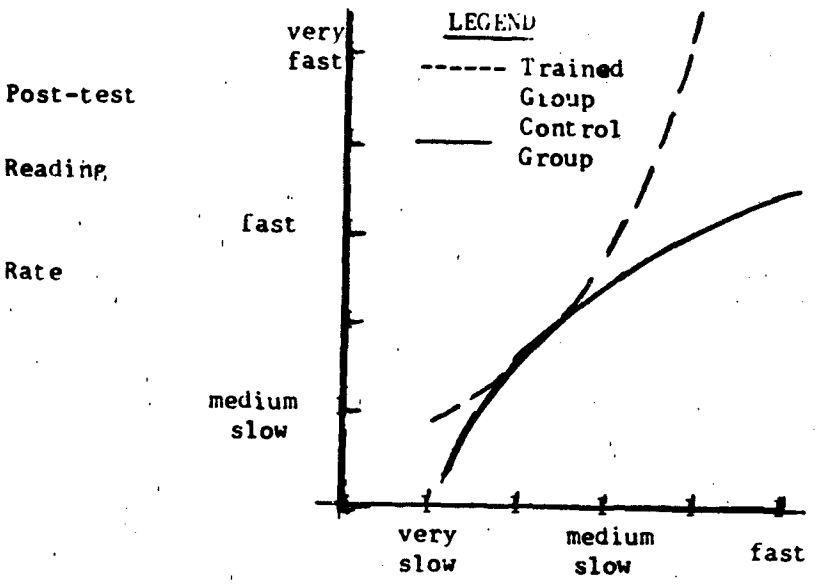


Figure 1. Theoretical curves showing expected change in reading rate for trained and untrained groups of adults. Absolute values depend upon the populations studied and the reading tests used.

Comparative Evaluation of CAI with Programmed Text and Lecture Modes of Instruction in Three Air Force Medical Training Courses

Gerard M. Deignan, Air Force Human Resources Laboratory
LtCol Seager, Maj Kimball, Capt Horowitz, Air Training Command

Abstract

The major objectives of this study were to compare computer-assisted instruction (CAI) with lecture and programmed text modes of instruction on dimensions of (1) instructional effectiveness, (2) time-savings, (3) attitudinal acceptance, and (4) instructional delivery costs. The first three dimensions are covered in the present paper. Precourse assessment measures were also developed to profile the characteristics (e.g., aptitude, biographical data and attitudes) of learners for whom CAI, lecture, and programmed text (PIT) might be differentially effective for medical tasks in segments of the three medical training courses. Degree of comparative instructional effectiveness and/or efficiency was found to vary chiefly as a function of (1) learner characteristics (e.g., aptitude and achievement motivation), (2) course related factors (e.g., difficulty), and (3) instructional mode.

Training sufficient numbers of skilled medical personnel in less time than costly conventional methods is of prime interest to medical administrators. Unfortunately, there is little empirical evidence to assist decision-makers in choosing among individualized, self-paced alternatives to lecture. Comparative analyses of the differential effectiveness of computer assisted instruction (CAI), programmed instructional text (PIT), and lecture methods of instruction in field settings have been sparse and sometimes equivocal (Dallman & DeLeo, 1977; Dare, 1975; Dwyer, 1972; Jamison & Suppes, 1974; Keesler AFB, 1974; McKeachie, 1974). Presently, we neither know whether one instructional method is more effetive than another for certain kinds of students confronted by different tasks, nor the degree of effectiveness. Rather, we behave as if (1) learners possess and employ to a similar degree the same characteristics for processing information and (2) instructional methods/media are equally effective for different kinds of tasks and students, despite the vast literature on individual differences and task difficulty.

To reduce this data gap, the present medical training study was conducted to provide answers to two major questions: (1) "Do computer assisted instruction (CAI), programmed instructional text (PIT), and lecture differ in instructional effectiveness in courses varying in difficulty?" and (2) "Do students who differ in characteristics (e.g., aptitude level and motivation) achieve more in less time under computer assisted instruction, programmed instructional text, or lecture?"

Method

Students. Three medical training courses (Medical Laboratory, Radiology, and Dental Assistant) were selected to provide a range of learner characteristics and course content suitable for generalizing results to students in medical courses of comparable difficulty. The Medical Lab course represented a difficult course requiring a relatively high aptitude

level. Radiology and Dental courses corresponded respectively to average and low difficulty courses of corresponding aptitude levels. The student sample during formative and summative evaluation consisted of 700 male and female trainees assigned to the Air Force School of Health Care Sciences at Sheppard Air Force Base, TX.

Pre-Course Measures. Based upon training task analyses, selected pre-course learner characteristics measures were developed and administered via automated slide-tape to all students prior to assignment to CAI, PIT, or lecture conditions. These measures included: (1) the medical version of the Delta Reading Vocabulary (r_{yy}=.88)(Deignan, 1973), (2) the General Aptitude Index from the Armed Services Vocational Aptitude Battery (ASVAB) (r_{yy}=.88), (3) the memory (r_{yy}=.66), and (4) visualization (r_{yy}=.77) and biographical measures from the Delta Training Aptitude Battery (Deignan, 1976). Measures were administered prior to course entry to aid CAI authors in the development and formative evaluation of instructional material appropriate to the target population in each course. These criterion-related control measures were subsequently used to assist summative evaluation in terms of explaining, interpreting, and generalizing comparative performance results.

Procedures. Within each course, CAI was compared to PIT in Radiology or Lecture in the Medical Lab and Dental courses. Criterion measures included post-instructional measures of achievement, elapsed time to complete instruction, and attitudes toward CAI, PIT, and Lecture. One hundred students were programmed for assignment to each CAI and non-CAI control condition in each course during summative evaluation. In some analyses, the sample size was less than 100 per condition because some students lacked complete data on pre-course assessment and/or criterion data.

Results

To provide evidence on the question of comparative instructional effectiveness independent of learner characteristics, main effects achievement and time to completion data are reported in Table 1.

Table 1
Mean Percentage Achievement and Mean Time in Minutes to Complete Instruction in three courses

Group	% Correct Achievement Score		Time to Complete Instruction (Min)	
	M	SD	M	SD
Medical Lab				
CAI N-93	88.94	13.51	460	230
Lecture N-96	75.12	13.84	540	0
Radiology				
CAI N-97	84.72	8.14	240	65
PIT N-89	81.95	10.62	271	106
Dental				
CAI N-101	77.03	13.59	453	95
Lecture N-52	78.07	14.01	540	0

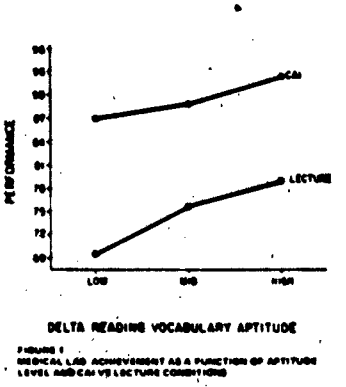
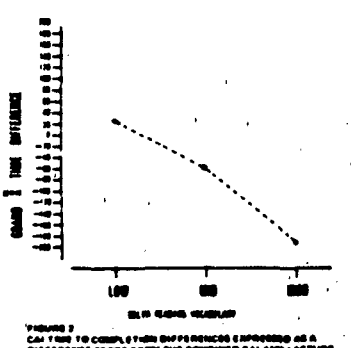


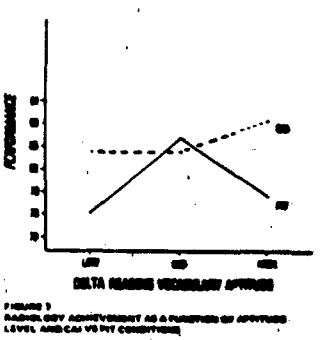
FIGURE 1
MEDICAL LAB ACHIEVEMENT AS A FUNCTION OF APTITUDE LEVEL AND CAI VS LECTURE CONDITIONS

CAI students demonstrated significantly greater achievement than lecture students in the Medical Lab course, $F(1,189)=50.91$, $p < .001$. Similarly, CAI students achieved more than their programmed text counterparts, $F(1,184)=4.32$, $p < .05$. Dental course achievement comparisons between CAI and lecture were not statistically significant. Comparisons of time to completion differences in Table 1 revealed: (1) Medical Lab CAI students completed instruction in 14% less time than Lecture students, (2) Radiology CAI students completed instruction in 12% less time than programmed text students, and (3) Dental CAI students required 17% less time than Lecture counterparts. Hence, in all three courses, CAI time to completion was significantly less than Lecture or self-paced programmed text when learner characteristics, e.g. aptitude were not considered.

To determine whether CAI, Lecture, or programmed text might be differentially effective for certain student subgroups, 2 (CAI vs Lecture or PIT) X 3 (Low-Middle-High Aptitude) analyses were performed. Figure 1 depicts Medical Lab achievement score differences as a function of aptitude level and CAI vs Lecture conditions. Analysis of variance of achievement scores revealed significant main effects for the CAI vs Lecture treatments, $F(1,177)=54.51$, $p < .001$ and aptitude level, $F(2,177)=5.41$, $p < .01$. Low aptitude CAI students scored 18 percentage points more than low aptitude lecture controls. Similar CAI achievement superiority was reflected at the middle and high aptitude levels. Thus, not only did low aptitude CAI students achieve significantly more than low aptitude lecture controls, their achievement exceeded high aptitude lecture controls.

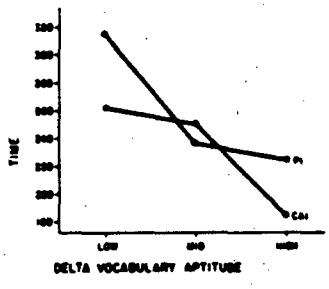
Medical Lab time to completion difference scores as a function of CAI-Lecture and aptitude level is shown in Figure 2.

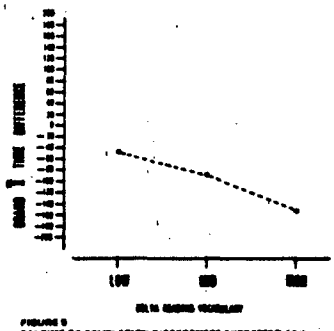




Medical Lab CAI students at the mid-aptitude level required 11% less time whereas high aptitude CAI students completed instruction in 33% less time. No statistically significant differences were obtained at the low aptitude level.

Comparison of Radiology achievement as a function of CAI-Programmed Text (PIT) and aptitude level is shown in Figure 3. In addition to a statistically significant aptitude effect, $F(2,149)=6.26$, p .01, a significant CAI-PIT treatment X aptitude interaction was obtained, $F(2,149)=9.22$, p .001. Inspection of the interaction shown in Figure 3 revealed both low and high aptitude CAI students scored 7-9% points higher than PIT students. Radiology time to completion analyses depicted in Figure 4 revealed significant aptitude, $F(2,149)=9.13$, p .001 and interaciton effects, $F(2,149)=6.69$, p .01. Low aptitude CAI students achieved more in 17% less time than controls. At the high aptitude level, CAI students time variability was 1/2 of the PIT students.





Dental achievement differences were not significant. However, a tendency did exist for low aptitude CAI students to achieve more than controls, despite criterion insensitivity. Significant Dental CAI completion time differences (Figure 5) ranged from 9% at the low aptitude level to 29% at the high aptitude level.

Discussion. In contrast to comparative treatment data reported in Table 1, learner characteristics data, e.g. aptitude, (Figs 1-5) provide valuable information and implications for training. For example, to optimize learner effectiveness and/or efficiency in the Medical Lab course, a scarce resource such as CAI might effectively be used to increase low

aptitude student achievement. To do so, however, means little or no time savings. If, on the other hand, achievement is not a problem, CAI may be assigned to Medical Lab and Dental course high aptitude students. Students similar to these are expected to complete instruction in 1/3 less time than if assigned to lectures.

Differences among courses and tasks, however, are important. Low aptitude CAI students within the Radiology course exceeded the achievement of low aptitude PIT students in 17% less time. CAI was no more effective or efficient than PIT at the mid-aptitude level. In contrast, CAI was demonstrably superior to lecture at all aptitude levels in the difficult Medical Lab course; whereas no achievement differences were obtained on the easier subject matter in the Dental course. Hence, data such as the preceding permit assignment of learners based upon learner characteristics to more effective or efficient instructional alternatives than merely assigning everyone to the same alternative.

Summary

To optimize the effectiveness and/or efficiency of instruction, comparative treatment data (Table 1) are necessary but insufficient. Learner characteristics by treatment information (Figs 1-5) provides additional data and implications for increasing learner effectiveness and/or efficiency. Moreover, optimizing instruction requires information pertaining to (1) course specific properties, (2) task difficulty and related learner characteristics, (3) student flow, (4) instructional alternative cost comparisons, and (5) tradeoffs regarding levels of instructional effectiveness vs time to complete instruction. Failure to obtain and consider such factors is a decision to maintain sub-optimal training effectiveness and/or efficiency when instructional alternatives exist.

REFERENCES

To conserve space, cited & additional references/data are included in: Deignan, G. M., Seager, B. R., Kimball & Horowitz, N. S. Computer-assisted, programmed text & lecture modes of instruction: Comparative evaluation in three medical training courses. AFHRL-TR-79-76, Lowry AFB, CO: Technical Training Division, Oct 1979.

Adaptations of Embedded Training

Lynn C. Percival
Honeywell Systems and Research Center
Minneapolis, MN 55413

Abstract

Embedded training (ET) refers to the use of a computer-based operational system for instructional purposes. The present paper describes ET concepts for a tactical sensor system and an automated test equipment system. It describes how ET can be adapted to meet personnel and system requirements. For the tactical system, ET capabilities include management and presentation of instruction, performance measurement and student performance record keeping. Functions include refresher and remedial training of individual and team skills. For the test system, ET capabilities include presentation of job-aids during test operations and presentation of instruction during dedicated training. Functions include cross-training, OJT, and refresher and remedial training. The reasons for including ET and its limitations are described.

Introduction

As our armed forces move toward the deployment of an increasing number of complex computer-based systems they are faced with a serious problem. Entry-level skills among military personnel continue to decline (Booher, 1979) while the increased capabilities of these computer-based systems require a high skill level among personnel to make use of these capabilities.

Embedded training (ET), which refers to training performed by an operational, computer-based system, is one method which has been proposed (Baker, 1978) to bridge the skill-capability gap. ET is a general concept of actual equipment being used to train the operator in its use. The present paper attempts to show how the ET concept can be adapted to fit divergent personnel and system requirements in two distinct systems (Percival, Ogden, Vomela, Kochevar and Brock, Note 1; Vestewig, Percival, Kochevar and Baum, Note 2; Levitan, Lysen and Prietula, Note 3). One of them is a Tactical Sensor System (TSS) which involves team operation, external communications, and a remote sensor. The other is an Automatic Electronic Test System (AETS) which involves a single maintenance operator and a Unit Under Test (UUT). The two systems have vastly different hardware and the operators in them perform tasks that have little or no overlap. What the two have in common is that they are computer based and have extensive capabilities that are likely to be under utilized.

Problem

Reasons for Including ET

TSS. ET must correct initial deficiencies in skill level. Typically operators of this system will be given several weeks of intensive training prior to assignment. Because of individual differences there will be a certain amount of skill level variation. In some cases, persons may be assigned to operate the system without the training but with skills that can be transferred under the appropriate conditions.

PRECEDING PAGE BLANK-NOT FILMED

Individual and team skills need to be exercised. Due to high operating costs, the system cannot be in operation a large part of the time. This means that the operational exercise of individual and team skills will be quite limited. Under these conditions the complex skills involved in operating the system are likely to decay or be forgotten. This decay will decrease system performance and operational readiness. It will also block personnel career advancement.

System maintenance skills require instructional practice. A large part of TSS maintenance is to be performed by the operator. However, without practice of these maintenance skills, such things as system maintenance logic, troubleshooting techniques, and use of software diagnostic routines are likely to be forgotten. This will have a negative impact on system readiness.

A hands-on component is a desirable addition to personnel career advancement testing. Previous testing has been largely of the paper and pencil variety. This probably does not represent an appropriate test of operator skills. As a result personnel career advancement may not have actually been based on job proficiency.

AETS. ET must minimize the effect of deficient skill level among electronic maintenance personnel. Deficient skill level may occur because the technician is fresh from school and has had no On-the-Job Training, (OJT) because the technician has no experience with this particular system, or because tests which are performed infrequently are subject to forgetting. In each of these cases the test system capabilities are not fully utilized.

ET should reduce the negative impact of late, hard-copy, test software (SW) documentation. Typically the delivery of new test SW leads its hard copy documentation by as much as one year. When this occurs the operator is at a severe disadvantage in using the new SW and the effectiveness of the test system is substantially reduced.

Inadequate or unsafe uses of the extended mode of operation should be reduced. In most cases test operation in the normal mode successfully isolates a fault. When it does not, the extended mode of operation is required. In the extended mode there is greater test flexibility for fault isolation, but this flexibility requires more advanced skill/knowledge from the operator. There is also an increased possibility of damaging the test SW, the system, or the UUT while in the extended mode.

ET must improve system maintenance. While the AETS will include self-diagnostic SW tests, problems remain in performing timely routine maintenance checks and in physically locating defective components isolated by the SW.

Limitations

Although there are several good reasons for including an ET capability in both the TSS and AETS there are several limitations that must be considered before implementing such a capability. The primary limitation in both is related to the time that the system will be available for training.

TSS. There are a number of persons in a TSS unit who require at least some training. This together with the requirement that operations be carried out part of the time means that training time per person is limited.

AETS. This type of system is utilized for test operation nearly all the time. The implication is that very little time is going to be available to dedicate the system for training.

Embedded Training Concepts

Capabilities

ET capabilities should be developed after consideration of needs, constraints, other instructional media, and the training pipeline for the system.

TSS. For this system there are four primary capabilities. These are: management of instruction, presentation of instruction, performance measurement, and student performance record keeping (Percival, et al, Note 1). These are described below.

The management function will direct all system specific training in the field by applying decision rules to current training doctrine, student baseline and performance history, refresher material schedule, and noted operational deficiencies. An effective management function results in efficient training by directing the student to the appropriate instructional media.

Presentation of instruction is limited to those operations which require hands-on interaction with the system hardware. Types of instruction include CAI lessons for individual operators and simulation of tactical operations at various team levels.

Measurement of individual and team performance will function in several ways. For the individual operator a skill profile can be developed that will aid instructional management. This same capability is required for personnel advancement testing. For the team, performance measurement will aid in assessing operational readiness.

The student performance record keeping capability is necessary for instructional management. A baseline and performance history data base is constructed that permits effective instructional management.

AETS. This sytem would have two primary ET functional capabilities. These are performance aiding during test operation and dedicated training (Vestewig et al, Note 2).

The aiding capability is an option that exists as part of a test during normal and extended modes of operation. There are several types of aids that can be presented. Decision aids (Levitan et al, Note 3) can be called to assist the operator during normal test operation. "Soft" limited documentation on recent test SW changes can be called up from normal or extended modes of operation. Maintenance aids consist of reminders for routine maintenance that appear as a function of time logged on the system. Graphics aids assist the operator in physically locating the isolated fault. In the extended mode of test operation the aids are designed to assist the operator in isolating faults that cannot be located during normal operation. These include such things as a summary of the test SW logic and a number of addresses where the results of intermediate tests not normally presented can be accessed.

Dedicated training on the AETS is confined to those critical areas in which the loss of system test capability is offset by the greater need to accomplish training. These include methods of repair for infrequent or uncommon UUT, station faults not trained in school or which require

updating practice, or major changes in diagnosis and repair procedures brought about by configuration changes.

Functions

TSS. ET functions for the tactical system include management of system specific training, delivery of refresher and remedial training at individual and team levels, operator maintenance training, exercise of individual and team skills, and individual, team, and system performance evaluation (Percival et al, Note 1).

AETS. For the test system ET functions include OJT, cross, refresher and remedial training using performance aiding, system maintenance aiding, and update and new procedural training in a dedicated training mode (Vestewig et al, Note 2).

Summary and Recommendations

The present paper sought to show how ET could be adapted to fit divergent systems. ET concepts for a tactical system and an automatic maintenance system were described. ET capabilities and functions were developed that took into consideration training needs, system constraints, the training pipeline, and other instructional media.

ET is a new, as yet untested, but extremely flexible concept that may be applied to a wide variety of systems. On the surface it is a cost-effective solution to a variety of training problems which continue to affect military systems. If it is to be used effectively experts in training, man-computer communication, job performance aiding, human factors, and military personnel systems need to be brought together to develop it so that it does help to reduce the gap between user skills and system capabilities.

Notes

1. Percival, L. C., Ogden, W. C., Vomela, R. W., Kochevar, J. W., and Brock, J. F. Technical report on embedded training concept for tactical target acquisition system. Minneapolis, MN: Honeywell Systems and Research Center, December 1979.

2. Vestewig, R. E., Percival, L. C., Kochevar, J. W., and Baum, D. R. Technical report on embedded training concept for automatic test equipment. Minneapolis, MN: Honeywell Systems and Research Center, December 1979.

3. Levitan, L. J., Lysen, P. E., and Prietula, M. J. Human factors for ATE. Minneapolis, MN: Honeywell Systems and Research Center, November 1979.

References

Baker, James D. Embedded training: The team component. Proceedings of the 22nd Annual Meeting of the Human Factors Society, 1978, 98. (Abstract)

Booher, H. R. Job performance aids research and technology base for integrated personnel systems. Proceedings of the 23rd Annual Meeting of the Human Factors Society, 1979, 483-486.

Comparison of Male vs Female Cadet Opinions Toward the USAF Academy Academic Program[1]

William H. Clover and David B. Porter

Abstract

Using survey data collected from over 2000 cadets in November 1978, attitudes of male and female cadets were compared. Results indicated strong consensus on dimensions and items about which cadets had strong feelings (i.e., faculty; scheduling flexibility; time pressure; and military decorum in the classroom). Differences between male and female attitudes were most pronounced on dimensions involving personal choice or preference (i.e., satisfaction with Academy; utility of math, science and engineering courses; importance of education; and use of instructor prerogative points). These differences generally showed that female cadets had slightly more positive attitudes toward academics especially in areas requiring verbal skills.

Extensive research has been conducted concerning attitudes and the integration of women into the Air Force Academy (DeFleur and Gillman, 1978a; DeFleur and Gillman, 1978b; Gillman and Harris, 1979). However, these papers have generally addressed male and female attitudes with respect to their socio-military roles at the Academy. To date there has been no systematic investigation of possible differences in male and female attitudes toward the academic program at the Air Force Academy. The purpose of this paper is to examine differences in male vs female attitudes toward the academic program.

METHOD

Development of Questionnaire. The Dean of Faculty requested a survey of cadet attitudes toward the USAFA academic program.[2] Cadet attitudes had never before been addressed in this manner. A key constraint was that the survey would be administered during classtime to help insure cooperation on the part of cadets. This constraint required the survey to be relatively short. The following steps were taken to generate an item pool: (a) literature on organizational and college "climates" was reviewed; (b) different faculty members were consulted; (c) approximately 30 cadets reviewed initial item pools for ideas and comments. This resulted in an item pool of approximately 150 items. The first author and several others

---

[1]Opinions expressed in this paper are solely those of the authors and do not necessarily reflect those of the Air Force Academy or the U.S. Air Force.

[2]Copies of this instrument are available from the authors upon request.

PRECEDING PAGE BLANK-NOT FILMED

reduced this set of items to the 67 most relevant items in several major areas.[3] Questions were asked using a 6 point Likert scale ranging from strongly disagree to strongly agree. There was no specific neutral point on the scale. An average value of 3.5 was considered neutral.

Administration and Sample. The survey was administered to all classes meeting during the third academic period on 11 Nov 78. This resulted in a total sample of 2145 (over 50% of the cadet population) cadets representing a broad cross-section of the cadet wing. A total of 1846 male and female responses was identifiable. Approximately 300 cadets did not provide the requested identification data. (The responses of this group were compared to those of the identifiable group and found to be nearly identical.)

Post Hoc Analysis of Questionnaire. The 67 item scale was submitted to factor analysis with varimax rotation using the statistical programs in SPSS (Nie, et al, 1975). Eleven factors were generated and items selected to belong to each factor (loading of .30 or greater) were subjected to an item analysis within each factor using item-total correlations. Items which did not correlate .50 or greater were dropped from that particular factor. Data on this final set of factors and selected individual items were presented to the Dean of Faculty and his staff.

Male and Female Sample Reported Here. Data from the freshmen (Class 82), sophomore (Class 81) and juniors (Class 80) are reported here. Seniors (Class 79) are excluded because there were no women in this class. It was decided the most meaningful data were those with the groups matched by class. Data reported here represent the opinions of 1389 males and 173 females.

RESULTS AND DISCUSSION

Table 1 reflects the mean responses of male and females on the different dimensions. The following paragraphs will highlight relevant distinctions.

Attitudes toward Faculty. Cadets who scored high on this dimension felt instructors were: knowledgeable; provided "quality" instruction; willing to give extra instruction; positive role models as Air Force officers; easy to know on a personal basis; and put a lot of energy and enthusiasm into their teaching. Both male and female cadets expressed very positive attitudes toward the faculty. This fact lends support to the argument that male and female cadets receive comparable treatment from faculty members. Finally, it should be noted that the dimension "Faculty" was rated as the most positive aspect of the academic environment. Additional support for this was noted in another paper by Clover and Porter (1980). The small standard deviations seem to indicate much more agreement among cadets on this issue than on others.

---

[3] It is ironic or an example of "functional fixidity" that no one thought to ask a question(s) dealing with "mandatory attendance at classes." This apparently was taken as a perceptual given.

Flexibility/Control of Academic Schedule. Cadets who scored high on this dimension typically wanted: more open academic options; to take more courses in their academic major rather than the required core curriculum;[4] a pass/fail option in core courses[5], and all physical education courses to be pass/fail.[6] Although both groups again expressed similarly strong positive attitudes toward increased flexibility and control over their academic schedules, the mean score for women (4.60) was slightly higher than the mean score for men (4.46). This is marginally significant (P=.06). This difference may be related to the preponderance of math, science, and engineering courses in the core curriculum and women's preference for Social Science and the Humanities (see Table 1, Item 10). Greater control and flexibility would allow cadets to take more courses they desire. Interestingly, there was a great divergence between male and female cadets on the question of whether or not Physical Education classes should be pass/fail. Female cadets indicated they preferred graded P.E. courses. One hypothesis for this result is that female cadets' attitudes were more favorable toward graded physical education classes because they expended more effort in this non-traditional area. Perhaps grades are perceived as more equitable compensation where a great deal of effort is involved.[7] This hypothesis raises several interesting conjectures concerning effort and motivation.

Acceptance of Standards/Personal Effort. Cadets scoring high on this dimension believe: in the academic standards at the Academy; and they can achieve high grades if they put forth the effort. Male cadets had significantly higher scores on this dimension than did female cadets. This result is interesting in light of the fact that females have somewhat higher entrance scores on their college boards than do males (Williams, 1979). However, academic performance of female cadets to date has only equaled or been slightly below that of males. This paradox between higher ability but lower performance might explain female cadets' lower average score on this dimension.

Satisfaction with the Academy. Cadets scoring high on this dimension like the Academy and would recommend it to a friend. Female cadets expressed significantly more satisfaction with the Academy in general than than did their male counterparts. This could be the result of post-decisional dissonance; female cadets are required to make proportionately

---

[4]The core curriculum consists of 37 course units, a typical academic major requires 11 additional course units with 2-3 optional classes. To take additional optional courses a cadet must either overload (take more than 6 course units in a given semester), receive transfer credit for previous courses, or be able to validate required core courses by examination.

[5]Presently only Military Training and Physical Education courses offer any core courses on a pass/fail basis.

[6]Some physical education courses are presently pass/fail.

[7]We wish to thank Major Dick Ellis of the Athletic Department for this insight.

Table 1

[illegible] Attitudes toward the Academic Program[1]

[illegible]	[illegible]	SD	Female	SD	t^2
1. [illegible]	4.21	.71	4.77	.58	
2. [illegible] of the Academic Program	4.44	.85	4.80	.51	.04
3. [illegible]	4.12	1.02	4.42	1.03	
4. [illegible]	4.27	1.05	4.65	1.04	*
5. [illegible]	4.17	1.00	4.52	1.16	*
6. [illegible]	4.74	.67	4.89	.75	*
7. [illegible]	3.93	.75	4.16	.72	
8. [illegible]	3.52	1.11	4.05	1.05	*
9. [illegible]	4.47	.79	4.76	.71	
10. [illegible]	4.17	1.00	4.64	.94	*
11. [illegible]	3.11	1.07	3.29	.99	

[1] [illegible]; the females = 173

[2] [illegible] at the .05 level.

[3] [illegible] the academic abilities [illegible] the attitudes [illegible] and [illegible] attitudes toward [illegible] and the abilities or traits.

greater sacrifices and based on the average higher entrance academic composite, perhaps forego greater alternative values. Only further investigation will reveal the reason for the differences.

Importance of Academic Education. Cadets scoring high on this dimension felt that: academic education was their main reason for coming to the Academy; academic accomplishments are important; and, both their major and their general academic education will be useful after graduation. Again female cadets expressed significantly more positive attitudes toward the importance of their academic education. A partial explanation of this is provided by the finding that significantly fewer female cadets "came to the Academy solely to fly." Another possibility is that many female cadets may plan to marry male cadets who will attend pilot training. Since the difficulties expected by a dual flying career family are greatly aggravated, it may be that many physically qualified female cadets have opted not to attend pilot or navigator training. Therefore, their academic education is perceived as being more pertinent when entering non-flying career fields.

Attitudes toward Academic Environment. Cadets scoring high on this dimension feel that: the curriculum and faculty encourage individualistic thinking and opposing viewpoints; the core curriculum is necessary; there are adequate formal feedback channels concerning instruction; that

instructors are knowledgeable about the non-academic aspects of cadet life; and instructors provide practical and realistic education. Male and female cadets expressed basically neutral attitudes toward this dimension. As with other dimensions which tapped perceptions of institutional reality (lack of flexibility and control over academic schedules and perceived time pressures), there was strong agreement by both sexes.

Attitudes toward use of "Instructor Points" (IP)[8]. Cadets scoring high on this dimension believe IP points should be a part of the course grade in both core and majors courses. Female cadets were significantly more positive toward this dimension than male cadets. Female cadets may be generally stronger in verbal skills, and instructor prerogative points reward this ability in the classroom. Since there was no significant difference in attitudes toward faculty, we do not feel that this reflects an expectation of faculty preference for female cadets but rather a recognition of female cadets' more highly developed verbal skills.

Alienation/Dissatisfaction with the Academic Program. In general, cadets who scored high on this dimension felt: there was too much emphasis on math, science, and engineering; that too much time was devoted to academics; that instructors did not understand the time pressures cadets were under; that grades received more emphasis than learning; and, that what they would be doing after graduation was more important than their academic performance. While there was no significant difference between males and females, it is interesting to note the trend by class. For the males the mean scores by year from freshman to senior year are 3.34, 3.58, 3.78, respectively. For the females, the mean scores from freshman through junior year are 3.28, 3.72 and 3.85 respectively. There does seem to be a trend for both males and females to become dissatisfied with the academic program. Only further study will determine the impact of this attitude.

Attitudes toward the Value/Utility of Math/Science/Engineering Courses Compared to Social Science/Humanities Courses. This scale turned out more ambiguous than others in its interpretation. Cadets who scored high i.e., greater than 4) seemed to favor math/science/engineering courses to the exclusion of social science/humanities courses. Cadets who scored low (i.e., less than 3) seemed to hold the opposite feelings. A neutral score (3.5) seemed to indicate a positive feeling toward both areas of study. Cadets who score high: feel there are not too many math science courses in the core; like "hard" sciences more than social sciences; believe officer tasks will require knowledge gained in math/science courses; and feel their most useful learning occurs in math/science/engineering courses. Cadets who score low: believe the major problems faced by officers will be "people" not "technical;" believe they can more easily apply the knowledge gained in the social sciences/humanities; believe they

---

[8] In contrast to objective measures such as test scores, instructor prerogative points are grades given to cadets based on the instructor's subjective evaluation of each cadet's participation in class discussions, oral presentations, preparation for class and/or understanding of the material. Those courses which use IP points as part of the grade typically limit them to a small percentage of the course grade, thus limiting their effect.

learn useful things in these courses; and believe there are not too many social science/humanities courses. In general, both groups scored low, however, the female cadets scored significantly lower than the male cadets. Attitudes of female cadets may well be caused by their greater ability in disciplines requiring verbal skills and their expectancy to enter predominantly non-technical career fields. The overall low level of these scores may indicate a general disenchantment with technical courses or simply be an artifact of this particular scale. If the former interpretation is accurate, than further study should be conducted to assess the impact of these attitudes.

Military Decorum in the Classroom. Cadets scoring high on this dimension believe that classroom military decorum[9] does not inhibit the learning atmosphere and contributes to their development as officers. There was little difference between male and female cadets' attitudes toward military decorum in the classroom. Both groups score below the neutral point. Apparently both groups have some reservation as to whether these behaviors in class contribute either to their development or their education.

Other Areas. Eighteen of the 67 questions did not correlate with the 11 dimensions and were treated separately. Several questions reflected close agreement in areas where cadets expressed strong opinions (i.e., mean scores greater than 4.5 or less than 2.5). For example, both male and female cadets felt that faculty members supported women cadets being at the Academy. They also agreed that the academic standards were higher at the Academy than at comparable civilian universities. Both groups felt that competition is emphasized more than learning, but that the material they learned in their academic major was more important than the grade they received.

Five questions reflected significant differences. Three questions related to academic orientation and two related to interest and motivation. Female cadets expressed stronger agreement compared to males that: working hard for grades is accepted by most cadets; Academic Call to Quarters is a reasonable requirement[10]; and, cadets need more time to study. Perhaps these differences add emphasis to female cadets' more positive attitude toward the importance of an academic education. A partial explanation of this is provided by their response to the statement, "I came to the Academy solely to fly when I graduate." The mean score for women was 2.17 compared to 3.07 for their male counterparts. Non-flying career fields may be

---

[9]Refers to such behaviors as: calling room to attention at the start and finish of each class; instructors addressing cadets either by "mister"/"miss" or by rank and last name; use of the military discipline to enforce academic performance (i.e., late homework, improper uniform, sleeping in class might incur appropriate military punishment).

[10]Academic Call to Quarters (known to cadets as the "Dean's Time") is approximately three hours on weeknight evenings which is set aside strictly for the purpose of academic study.

perceived as more "academic" than flying careers and since more women plan to enter these non-flying career fields, they feel academics are both more important and useful. Women also expressed stronger agreement (4.15) that social science and humanities majors have less status at the Academy than cadets majoring in engineering and science than did male cadets (X=3.66). This may reflect the perception that science and engineering are more closely related to flying duties and thus cadets are afforded higher status in the Cadet Wing if they are "technical" majors.

CONCLUSION

While there are many similarities between the attitudes of men and women, there are some important differences. Only future research will answer some intriguing questions raised by these data and their implications for the academic program at the Air Force Academy.

REFERENCES

Clover, W. H. and Porter, D. B., "Cadet Opinions of the Best and Worst Aspects of the USAFA Academic Program," Proceedings of the 7th Psychology in the DOD Symposium, USAF Academy, April 1980.

Defleur, Lois B. and Gillman, David C., Sex Integration of the U.S. Air Force Academy: Changing Roles for Women. Armed Forces and Society, Vol 4:4 (August '78)(a)

Defleur, Lois B. and Gillman, David C., Cadet beliefs, attitudes and interactions during the early phases of sex-integration. Youth and Society, Vol 10:2, 1978, 165-190.

Gillman, David C. and Harris, Dickie A., An Evaluation of Anxiety and Stress during Combat-Oriented Survival, Resistance, and Escape (SERE) Training: A Comparison of Female and Male Air Force Academy Cadets. Frank J. Seiler Technical Report 79-0007. Air Force Systems Command, Sep 1979.

Nie, N. H.; Hull, C. H.; Jenkins, J. G.; Steinbrenner, K. and Bent, D. H. Statistical Package for the Social Sciences (2nd Ed). New York: McGraw Hill Book Co, 1975.

Williams, J. W., "The Integration of Women Cadets into the Air Force Academy" Proceedings: 1979 American Psychological Association meeting, New York, September 1979.

Table 1

Mean Scores of Male and Female Cadets on Various Dimensions Relating to Attitudes toward the Academic Program[1]

Dimension	Male	SD	Female	SD	Sig[2]
1. Attitude toward Faculty	4.71	.61	4.72	.58	
2. Flexibility/Control of Academic Schedule	4.46	.85	4.60	.81	.06
3. Perceived Time Pressure	4.32	1.02	4.42	1.03	
4. Acceptance of Standards/ Personal Effort	4.27	1.05	4.03	1.04	*
5. Satisfaction with the Academy	4.09	1.30	4.32	1.18	*
6. Importance of Academic Education	4.04	.93	4.39	.75	*
7. Attitudes toward Academic Environment	3.62	.75	3.63	.72	
8. Attitudes toward Use of Instructor Points (IP)	3.52	1.41	4.05	1.35	*
9. Alienation/Dissatisfaction with the Academic Program	3.47	.78	3.46	.76	
10. Attitudes toward Value/Utility of Math, Science and Engineering Compared to Social Sciences and Humanities[3]	3.17	1.00	2.64	.94	*
11. Decorum in the Classroom	3.11	1.07	3.18	.99	

[1] N for males = 1389; N for females = 173

[2] *Significant at the .05 level.

[3] A high score on this dimension indicates a positive attitude towards the value of math/science/engineering classes. A low score indicates a positive attitude toward social science and humanities courses.

VISUAL SEARCH IN TWO-REGION CLUTTER DISPLAYS[1]

Jerry Wald, John R. Bloomfield,
Laura A. Thompson, Deborah A. Smith

Honeywell Systems and Research Center
2600 Ridgway Parkway
Minneapolis, Minnesota 55413

Search displays with well-defined regions containing high, medium, or low clutter densities were used in three experiments. In all three experiments, observers had to locate a target differing in color mixture from the clutter stimuli. In Experiment 1, four observers compared one-region displays with two-region displays of similar clutter density when both regions were searched. In Experiment 2 the same displays were searched by four observers who knew which region of the two-region displays contained the target. In Experiment 3, six observers saw two-region displays with different clutter densities in each region. Search time was influenced by clutter density, number of regions searched, and the presence of a neighboring nonsearched area.

Visual target acquisition performance has been shown to be influenced by the number and/or density of the nontarget clutter objects that are around the target. In general, it has been shown that the more nontargets that are present, the longer it takes to find a target (for example, Smith, Kincaid, and Semmelroth, 1962), although a reversal of this effect occurs as clutter density is increased to its limit (Bloomfield, Beckwith, Emerick, and Tei, 1975). In these experiments, search typically occurs in a single well-defined search area. What has not been studied is the effect of nontarget number/density on target search when there is more than one well-defined area to be searched or when there is a nonsearched area adjacent to an area being searched. The three experiments reported here address these issues. Experiment 1 investigates the effect on target acquisition performance of searching one vs two areas having the same nontarget density. Experiment 2 examines the effect of a nonsearched area on the search of an adjacent area having the same nontarget density. Experiment 3 explores the effect of having adjacent nonsearched areas that differ in nontarget density from the area being searched.

---

[1]This research has been supported by the U.S. Army Mobility Equipment Research and Development Command (MERADCOM) under contract number DAAK 70-79-C-0032. We wish to thank Dr. Kenneth Oscar of MERADCOM for his advice and encouragement.

METHOD

Subjects. Four observers participated in Experiment 1, four in Experiment 2, and six in Experiment 3. Each had normal color vision and a Snellen acuity (in some cases with correction) of 20/20 or better. One observer in Experiment 2 and two in Experiment 3 had previous experience with the task. All observers were between 16 and 35 years of age and were paid for their participation.

Stimuli. Commercially available vinyl asbestos floor tiles were cut into 2.54 cm squares and backed with magnetic tape. Two types of tile were used. They were similarly patterned with small patches of color but differed in the range of colors they contained. For convenience, they are referred to by the name of their dominant hue--yellow or white. (For a specification of these tiles in terms of the Munsell Color System, see Bloomfield, 1979.)

"Double" (54.62 cm x 81.93 cm) and "single" (54.62 cm x 41.28 cm) aluminum display boards were painted with gray magnetic paint. Yellow tile squares were placed in 47.00 cm x 36.20 cm stimulus areas. These squares were arranged in 17 x 13, 9 x 7, or 5 x 4 rectangular arrays to create high, medium, or low nontarget clutter density areas. Each single board had one such area. Each double board had two such areas with a gap of 4.45 cm between them.

In Experiments 1 and 2, single and double boards of each clutter density were used. In Experiment 3, double boards with a medium clutter density in one of the two search areas and either a high or a low clutter density in the other were used.

Procedure. In all three experiments, observers participated individually in experimental sessions that lasted approximately 1.5 hours.

In Experiments 1 and 2, each observer searched six display boards in each experimental session and participated in eight to 12 sessions. Each display board represented one of the six factorial combinations of display board (single, double) and nontarget clutter density (low, medium, high). In Experiment 1, both areas of the double boards were searched. In Experiment 2, search was focused on one side only. Left and right sides of the double boards were focused on alternately across sessions and counterbalanced across observers.

In both Experiments 1 and 2, there was one block of ten trials for each board per session. A different 6 x 6 Latin square was constructed for each observer to determine his/her board order during the first six sessions. In subsequent sessions the orders were repeated cyclically, as necessary. Across the board orders of each Latin square, every board preceded and followed every other board only once.

In Experiment 3, the observers searched the medium-density side of the two display boards. In each session there was a block of 31 trials with one board and a block of 32 trials with the other. The number of trials per board alternated between boards on successive days so that, every two sessions, each of the 63 possible target positions on the medium-density side of each display was sampled once.

For three of the six observers the medium-density side of the board always appeared on the right side of the two displays; for the other three

observers it always appeared on the left. Board order was balanced across observers, across days.

Before each trial in all three experiments, a white square, the target, replaced a randomly selected yellow nontarget on a display. Positions were selected with replacement in Experiments 1 and 2 and without replacement in Experiment 3. The observer sat in a darkened room facing a two-way mirror oriented at a 45-deg angle to his/her line of sight. The observer initiated a trial by pressing a switch that 1) turned off the lamps lighting a blank field reflected into the mirror, 2) turned on lamps for the display on the other side of the mirror at the same level of illumination, and 3) started an electric timer. The observer terminated the trial by releasing the switch, which reversed the lighting and stopped the timer; he/she then reported the target's position. The time required to find the target was recorded and reported to the observer after each trial.

RESULTS AND DISCUSSION

The data from the last six sessions of each of the three experiments were plotted as cumulative search time distributions. One subject's data from Experiment 1 are shown in Figure 1 as an example. There were large individual differences among observers in all three experiments.

In Experiments 1 and 2, each observer's cumulative distributions were used in nine Mann-Whitney U tests: three clutter density comparisons were made for single-board data (low vs medium, medium vs high, low vs high), three analogous comparisons for double-board data, and three single- vs double-board comparisons, one for each clutter density. The results of these comparisons for each observer are shown in Table 1.

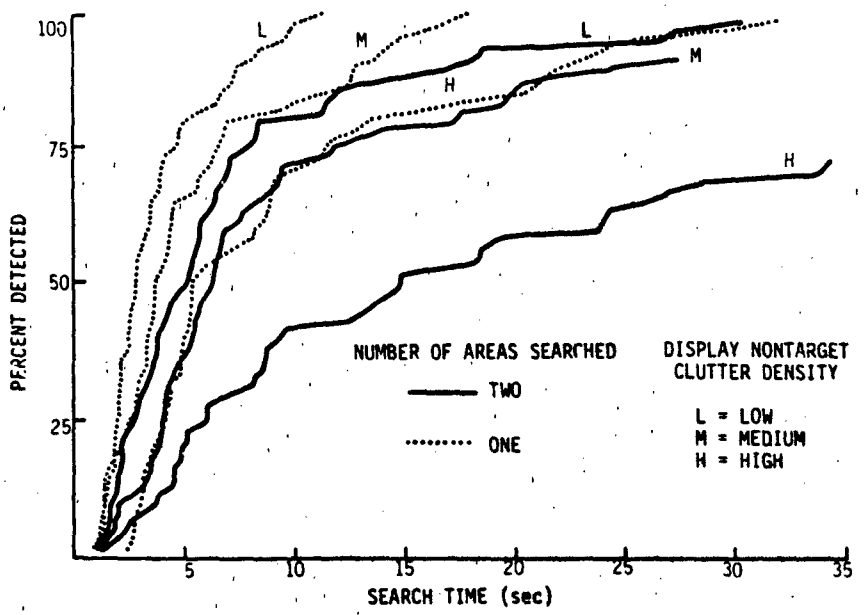


Figure 1. Cumulative Search Time Distributions Obtained When Observer Searched Displays with One or Two Search Areas

TABLE 1. SIGNIFICANCE LEVELS OF MANN-WHITNEY U TESTS FOR CLUTTER DENSITY AND DISPLAY BOARD COMPARISONS*

Observer	Within Display Comparisons						Between Display Comparisons		
	Single-Board Displays			Double-Board Displays			Single- vs Double-Board Displays		
	L vs M	M vs H	L vs H	L vs M	M vs H	L vs H	L vs L	M vs M	H vs H
Experiment 1									
1	0.05	0.001	0.001	0.05	0.001	0.001	0.001	0.001	0.001
2	0.05	NS	0.001	NS	0.05	0.01	0.001	0.05	0.01
3	0.001	0.05	0.001	0.001	0.01	0.001	0.001	0.001	0.001
4	0.05	0.001	0.001	0.01	0.001	0.001	0.01	0.001	0.001
Experiment 2									
1	0.001	NS	0.001	0.001	NS	0.001	0.05	NS	NS
2	0.001	0.05	0.001	0.001	0.01	0.001	0.01	NS	NS
3	0.001	0.001	0.001	0.001	NS	0.001	NS	0.05	NS
4	0.001	NS	0.001	0.01	NS	0.01	0.05	NS	NS

*L = Low nontarget clutter density
M = Medium nontarget clutter density
H = High nontarget clutter density

NS = Not significant

In Experiment 1, the observers searched either one or two areas of a cluttered display. Of the 24 clutter density comparisons (for single and double boards) shown in Table 1, 22 showed that the lower of the two clutter densities produced a significantly faster search time distribution ($p < 0.05$); one medium-vs-high comparison for the single boards and one low-vs-medium comparison for the double boards were not significant. In the 12 single- vs double-board comparisons, the single boards always produced significantly faster search time distributions ($p < 0.05$). Thus both clutter density and number of areas searched had a marked effect on the observers' search time distributions.

In Experiment 2, the observers searched only single areas, whether on a single- or double-board display. Again, of the 24 clutter density comparisons for single and double boards shown in Table 1, 19 showed that search time densities for the lower of the two clutter densities were significantly faster than those for the higher ($p < 0.05$). The other five comparisons were not significant; they involved medium vs high clutter density comparisons--two for the single boards and three for the double boards. In the 12 display board comparisons the search time distributions for the single boards were significantly faster ($p < 0.05$) in three of the four low-density comparisons and in one of the four medium-density comparisons. The other eight comparisons, including all four high-density comparisons, were not significant.

The results of this experiment indicate, again, that clutter density affects search time distributions. The results also indicate that the presence of a nonsearched area adjacent to a searched area can affect search performance, particularly when low-density areas are involved.

In Experiment 3, medium-density areas of double boards were searched. These areas were adjacent to either a high- or low-density area that was not searched. The search time distributions obtained for each observer on these two displays were compared, again using the Mann-Whitney U Test. The density of nonsearched area significantly ($p < 0.05$) influenced search performance for only one of the six observers, the low-density area producing significantly faster search times. This failure to obtain a density effect is consistent with the finding of Experiment 2--that search of single areas on medium- and high-density double-board displays is not greatly affected by the presence of a nonsearched area with the same clutter density. It seems likely that the particularly pronounced adjacency effect of the low-density double board in Experiment 2 is attributable to its Gestalt properties, which differ from those of the medium- and high-density double boards. This possibility will be explored in further experiments.

REFERENCES

Bloomfield, J.R. Visual search with embedded targets: Color and texture differences. Human Factors, 21, 317-330, 1979.

Bloomfield, J.R., Beckwith, W.E., Emerick, J., and Tei, B.E. Visual search with complex stimuli: Number and separation effects in density experiments. Paper presented at Human Factors Society 19th Annual Meeting, Dallas, Texas, 1975.

Smith, S.W., Kincaid, M.W., and Semmelroth, C. Speed of visual target detection as a function of the density of confusion elements (Memorandum 2900-235-R). Institute of Science and Technology, University of Michigan, 1962.

Brain Lateralization and Simple vs Integrated Task Performance[1]

Linda D. Jackson, Richard L. Hughes and Gene A. Berry
United States Air Force Academy

Abstract

Forty U. S. Air Force Academy cadets, grouped by sex and dominant hand, were compared on spatial and sequential tasks performed both independently and jointly. Prior research has suggested that each task is controlled in a different brain hemisphere. When the two tasks were performed independently there was no handedness effect. However, when both tasks were performed simultaneously, there was a significant advantage for right-handers ($p < .025$). There was also a slight advantage for males on both the independent spatial ($p < .05$) and joint task ($p < .10$). Results were attributed to the hemispheric interference left-handers experienced on the joint task due to their less distinct spatial-sequential hemispheric specialization. The preponderance of right-handers is posed as possibly resulting from the evolutionary advantage of hemispheric lateralization for such integrated tasks.

An old question is "Why are most persons right-handed?" Many answers have been posed, but a recent line of research may possibly contribute to the answer. Many studies (Diamond & Beaumont, 1974; Gazzaniga, 1967) have described abilities for the right and left hemispheres including the sequential/verbal superiority of the left hemisphere and spatial superiority of the right hemisphere. Bilateralization, or lack of specialization, is considered to impose some relative deficit in cognitive functioning. For example, a relative absence of brain lateralization has been used to explain sex-related cognitive differences (Levy, 1972) and also a presumed cognitive deficit for left-handers (Levy-Agresti & Sperry, 1968). Levy hypothesized that the inferiority of females on spatial tasks is due to verbal bilateralization, which interferes with right hemispheric spatial functioning. Left-handers, too, more frequently have bilateral sequential/verbal representation, although they do not necessarily show a deficit on spatial tasks (Kinsbourne, 1979). The present study explored sex and handedness as variables affecting performance on tasks apparently mediated primarily by the right or left hemispheres.

Handedness would seem to reflect some degree of brain lateralization. Clearly, most humans are right-handed, and the strength and consistency of right-side bias probably has a biological basis. It seems reasonable that this right-side bias may be related to the dominance of the left hemisphere in cerebral functioning, which is also characteristic in humans. This is consistent with the neurological fact that the right and left cerebral hemispheres exert a contra-lateral control over the body with the left hemisphere controlling the right side and vice versa. Kinsbourne (1979) speculated that this arrangement has adaptive significance in giving the same side of the brain control over the coordinated acts of pointing and naming. Inconsistent with this view, however, is the mass of data demonstrating that left-handers are just as proficient on cognitive tasks as right-handers (Hardyck, Petrinovich, & Goldman, 1976). Some deficit would be expected among left-handers if brain lateralization were advantageous

PRECEDING PAGE BLANK-NOT FILMED

since left-handers are typically less well lateralized than right-handers (Kinsbourne, 1979).

The inability of investigators to document a cognitive deficit in left-handers raises questions about any possible functional advantage of hemispheric specialization. Furthermore, it seems inconsistent with the right-side bias, a bias which suggests some evolutionary advantage. One hypothesis which may reconcile these apparent discrepancies was offered by Kinsbourne (1979). Kinsbourne reasoned that the advantage of hemispheric specialization may only be observed when an individual is performing two tasks at the same time. Left-handers may well indeed perform more poorly than right-handers when performing two simultaneous tasks.

Kinsbourne's logic was that if brain control centers are required to mediate different actions at the same time, there is less interference if the centers are poorly connected. If left-handers fare more poorly than right-handers on combined two-hemisphere tasks, that would provide some indication that there is a functional advantage to hemispheric specialization. It would suggest that previous studies failed to detect the deficit in left-handers because those studies only examined such tasks performed individually.

Brain lateralization is also used to explain frequently-observed sex differences in cognitive functioning, especially early female verbal superiority. Levy (1972) likened females to left-handed males, hypothesizing that both groups were more bilateral with regard to verbal function. This presumed verbal bilaterality would then interfere with right-side spatial function.

Therefore, if there are indeed differences in bilateralization based on both sex and handedness, it should be demonstrated via performance on integrated spatial and verbal-sequential tasks rather than on simple, independent tasks. The purpose of this study was to test this hypothesis of the effects of brain lateralization on the performance of simultaneous tasks which in lateralized subjects are controlled by opposite hemispheres.

Method

Subjects

Spatial and sequential tasks were performed both independently and simultaneously by forty volunteer college students assigned to four groups of ten based on handedness and sex. Handedness was determined using the Edinburgh Handedness Inventory (Oldfield, 1971). All subjects performed the independent tasks first, then the two tasks simultaneously. One half of the subjects performed the spatial task then the sequential task, the others in the reverse order.

Procedure

Spatial task. The independent spatial task involved replicating an irregularly shaped design portrayed on a slide. The designs were similar to those used in the WAIS Block Design sub-test, although more complicated. Subjects constructed the design using two inch square ceramic tiles, divided diagonally into red and white halves. Subjects first practiced the task by duplicating a 3 X 3 tile design and then a 5 X 5 tile design. The measured spatial task was a second 5 X 5 tile design. Performance was measured as the number of seconds required to correctly duplicate the design using the tiles.

Sequential task. The independent sequential task required the subject to respond to a taped list of sequentially descending (by fourteens) four digit numbers by subtracting seven and verbalizing the result aloud before the next taped number was presented. Each subject responded to 25 numbers and the score was the percentage of correct responses.

Simultaneous Spatial-Sequential task. Following performance on both independent tasks, each subject performed different but comparable tasks simultaneously. In other words, while attempting to duplicate a 5 X 5 tile figure, the subject also responded to a taped set of descending four digit numbers by subtracting seven from each. For this simultaneous task there were two dependent measures - number of seconds to figure completion and percentage of correct number responses during this period. Either task constituted a performance limitation on the other.

Data Analysis

To answer the questions posed in this study, three analyses were performed. Both independent variables (sex and handedness) were compared on each independent (spatial and sequential) task. A multivariate analysis of variance was performed for the simultaneous spatial-sequential task, jointly assessing effects on the two dependent variables. Those variables were seconds to figure completion and percentage of correct number responses.

Results

An analysis of variance was performed for each independent task. The resulting Fs for spatial task performance indicate a sex difference favoring males [F (1,36) = 5.02, $p < .05$], no handedness effect [F (1,36) = .56], and no interaction [F (1,36) = .93] between the two variables. For the independent sequential task there were no performance differences between sexes [F (1,36) = .23] or dominant hands [F (1,36) = .23], and again no interaction [F (1,36) = .39]. These analyses produced the expected slight male superiority on the purely spatial task, no sex difference on the sequential task, and neither dominant hand differences or sex-hand interactions when either task was performed independently.

A multiple analysis of variance for the combined spatial-sequential task produced quite different results. Fs converted from Wilk's Lamdas with df = 2 and 70 are as follows: $F_{Hand} = 4.14$, $p < .025$; $F_{Sex} = 2.38$, $p < .10$; $F_{HxS} = 1.73$, nonsignificant. These results indicate that when the two types of task were performed simultaneously, there was some male superiority, a much clearer performance advantage for right-handers, and no apparent sex-hand interaction.

Discussion

In this study groups categorized by handedness and sex performed different simple tasks both independently and simultaneously. The two tasks were ones supposedly controlled by different hemispheres of the brain. If performance of left vs. right-handers and males vs. females were comparable when the tasks were performed independently, but different when performed simultaneously, then it would indicate a difference in degree of hemispheric specialization as reflected by differential interference between the tasks.

In addition to the expected slight sex differences on the spatial task, there were substantial performance differences favoring right-handers on

the simultaneous tasks. Previous research which has failed to demonstrate any cognitive deficit in left-handers probably did not pose the correct question. Kinsbourne's (1979) view that left-handers might be more handicapped when performing two different tasks at the same time has been supported by these results. While there were no differences between right and left-handers when the tasks were performed separately, left-handers did perform significantly more poorly than right-handers when the tasks were performed simultaneously. Apparently, the lesser hemispheric specialization characteristic of left-handers did contribute to greater interference between the two tasks, which were obviously designed to favor asymmetry. The results of this study support the view that hemispheric lateralization developed because it offered an evolutionary advantage on multiple task performance. One effect of lateralization may be the preponderance of right-handers.

References

Diamond, S. J. & Beaumont, J. F. (Eds.) Hemisphere function in the human brain, New York: Wiley, 1974.

Gazzaniga, M. S. The split brain in man. Scientific American, 1967, 217, 24-29.

Hardyck, C. Petrinovich, L. F., & Goldman, R. D. Left-handedness and cognitive deficit. Cortex, 1976, 12, 266-279.

Kimura, D. The asymmetry of the human brain. Scientific American, 1973, 228, 70-78.

Kinsbourne, M. Why is the brain biased? Psychology Today, 1979, 13, 150.

Levy, J. Lateral specialization of the human brain: Behavioral manifestation and possible evolutionary basis. In J. A. Kiger (Ed.) The biology of behavior. Corvalis: Oregon University Press, 1972.

Levy-Agresti, J. & Speery, R. W. Differential perceptual capacities in major and minor hemispheres. Paper presented at fall meeting, National Academy of Sciences, California Institute of Technology, Pasadena, Proceedings of the National Academy of Science, 1968, 61, 1151.

Oldfield, R. C. The assessment and analysis of handedness: The Edinburgh Inventory. Neuropsychologia, 1971, 9, 97-113.

Footnote

[1]The views expressed herein are those of the authors and do not necessarily reflect the views of the United States Air Force or the Department of Defense.

THE EFFECTS OF REDUNDANT MULTIDIMENSIONAL DISPLAY CODES IN A COMPLEX INFORMATION PROCESSING TASK

Richard E. Christ and Gregory M. Corso[1]
New Mexico State University

Comparisons were made between the use of color and shape coding in a relatively complex multiple task situation. No differences in performance were attributable to the effects of these two visual codes. The use of shape and/or color in a multidimensional coding format did have effects on performance, but the direction and magnitude of those effects were a function of the task required and the dependent measure employed.

Because of technological advances in visual display design, the question of the effectiveness of color as a visual code is of great current interest. Previous work in our laboratory has evaluated the effectiveness of color coding relative to achromatic letters, digits, and shapes. The basic approach was a comparison of the effects on simple task performance of presenting information by these four codes. The simple tasks of choice reaction, search and locate, and identification-memory were performed singly or in irregularly alternating combination. The results of those previous studies provided no basis for concluding that color has any particular advantage or disadvantage to task performance that makes it different from the achromatic codes used for comparison.

The conclusions from those previous studies must be qualified by the fact that they apply only to simple discrete tasks. The laboratory work reported here employed a more realistic representation of an operational situation; one that gives the simple tasks meaning within a more complex set of continuous activities. The subject was required to monitor and effect changes in the parameters of aircraft presented in an air traffic control problem and, in addition, to respond as requested to the same types of simple tasks as used previously, but now as subtasks in the context of the more complex continuous task. Once the subject had learned to perform this air traffic control problem with considerable proficiency, two other coding variables were introduced to represent two different classes of information. First, the approximate altitude of each controllable aircraft was dichotomized and either shape coded or color coded. Secondly, both the approximate altitude and the approximate speed of each controllable aircraft were encoded; shape was used to encode altitude and color was used to encode speed or vice versa. The question in this experiment was whether or not the shape and color coding of two major aircraft flight parameters would (a) aid the subject in his attempt to monitor and control the aircraft and (b) aid the subject in his attempt to respond to the discrete tasks. The latter issue was based on the argument that shape and/or color coding would reduce the stress inherent in the air traffic control problem and consequently increase the subject's ability to respond to the discrete tasks.

---

[1]G.M. Corso is now at the Georgia Institute of Technology

Method

Since the apparatus used for this experiment has been explained in detail in previous reports, only the essential features will be described here.

A central display consisting of a 4 x 4 matrix was designed to represent an airspace on to which anywhere from two to eight coded aircraft could be rear projected. The coded aircraft could be repositioned in discrete steps to represent the movement of the aircraft from one location to another. The stimuli used to encode aircraft were drawn from a population of six different letter-digit combinations. In later phases of the experiment those stimuli which were under the subject's control (C2, C4, N4, and N7) were circumscribed by either an open circle or an open square and the two shapes were either white or were color coded (red or green). These controllable "aircraft" could appear at most only once each in the central display. The other two stimuli (P2 and P7) represented aircraft which were not controllable by the subject and which could occur up to four times each in the central display.

The flight parameters under direct control of the subject were altitude, speed, and heading. Detailed quantitative information concerning an aircraft's controllable flight parameters could be requested by the subject; this information was presented in digital form in three peripheral displays. Change in any one of these three parameters could be made through a numeric keyboard. The subject's task was to prevent any controllable aircraft from leaving the display or from crashing to the ground. The latter would occur if the aircraft exceeded acceptable ranges of speed and altitude or if two controllable aircraft occupied the same cell of the matrix with a vertical separation of less than 500 feet.

Displays which signalled the occurrence of a discrete task were located in different positions around the periphery of the display console. The comparison task required the subject to indicate whether or not three peripheral displays contained the same aircraft code. The choice reaction task required the subject to identify which one of six equally likely aircraft codes had been presented. A search and locate task required the subject to find a designated aircraft code in the central display and to indicate in which of the four quadrants the signal was located. An identification-memory task required the subject to give either a partial or a full-report of all the stimuli in the central display at the time the discrete task was presented. The single task displays were presented at random intervals and remained on for a maximum of three seconds or until the subject made an appropriate response. The status of stimuli in the central display was frozen in time during the discrete task response interval but, except for the identification-memory task, the central display was presented continuously. The central display was terminated when the identification-memory display was presented.

Six highly practiced subjects from our previous studies were used in this study. They were paid a fixed sum and also a monetary incentive designed to maximize their performance levels.

Experimental Design and Procedure

The aircraft control task was designed to be five minutes in duration. The number of aircraft symbols and their initial location, heading, speed, and altitude were predetermined and balanced over successive trials. There were from two to four controllable aircraft symbols and from zero to four uncontrollable symbols.

After receiving considerable practice with the new target codes and in controlling the parameters of the aircraft, each subject was run through three successive blocks of experimental sessions. Three sessions, each with seven five-minute trials, constituted a control phase (Phase 0). Then, all subjects were introduced to the altitude encoding in the central display (Phase 1); three subjects had color coded altitude information and three had shape coded altitudes. Three training sessions, as described for the control phase, were followed by a fourth session in which altitude encoding occurred on four criterion trials followed by four control trials. In the third block of sessions (Phase 2) both altitude and speed were encoded in the central display. Again three training sessions were followed by a fourth session containing four criterion followed by four control trials.

Results and Conclusions

The only measure of performance analyzed from the aircraft control problem was the count of the number of aircraft flying at the end of each five-minute run. These data were summed over the seven training runs in each of Sessions 1-3 for each phase of the experiment and over the four criterion and four control runs in the fourth session of Phases 1 and 2. There were three measures of performance for the choice reaction, search and locate, and comparison tasks, each pooled over the three times they were requested per run and then pooled over successive runs per session or partial session: percentage of trials to which the subject responded, the percentage of responses which were correct, and the mean correct response time. There also were three dependent measures for the identification-memory task, pooled over successive runs per session separately for the full- and partial-report conditions: mean number of correct reports per run, the mean time to make the first response per run, and the overall mean response rate per run.

The results of the study can be summed up simply: with practice on the task to the point where subjects were operating the air traffic control task proficiently, there were no differences in performance on the air traffic control task or on the simple discrete task that could be related to a difference between the effects of shape and color coding.

There was evidence that the multidimensional coding did affect performance. In Phase 1 of the experiment, the subjects performed better in the aircraft control problem during the control trials than in the trials which used color or shape to encode altitude; there was no difference in performance between control trials and the criterion trials in Phase 2. There was no effect of multidimensional coding on the accuracy data derived from the choice reaction, search and locate, or the comparison tasks. Correct response time in the search and locate task and in the comparison task were

shorter in the trials which used color or shape to encode altitude than in the control trials. There was no difference in response time between control trials and criterion trials when the latter involved the encoding of both altitude and speed and there was no effect of shape and/or color coding in the choice reaction task. Shape and/or color coding reduced the accuracy of responses in the identification-memory tasks but had no effect on overall response rate. First-response time in the multiple response identification-memory task was affected by the shape or color code but only when a partial report was required and then only when altitude was encoded. There was no effect on first response time in the full-report condition and the encoding of both altitude and speed (with shapes and color) produced an interference effect which offset the benefits of practice.

In conclusion, there appeared to be no differences in performance that could be attributed to a difference between the effects of shape and color coding. However, the use of shape and/or color in a multidimensional coding format did have effects on both continuous and discrete task performance, but the direction and magnitude of those effects were a function of the task required and the dependent measure employed.

References

Christ, R. E. Review and analysis of color coding research for visual displays. Human Factors, 1975, 17, 542-570.

Christ, R. E., & Corso, G. M. Color research for visual display (ONR Report CR213-102-3). Las Cruces, New Mexico: New Mexico State University, Department of Psychology, July 1975.

Teichner, W. H., Christ, R. E., & Corso, G. M. Color research for visual displays (ONR-CR213-102-4F). Las Cruces, New Mexico: New Mexico State University, Department of Psychology, June 1977.

Individual Differences and Laterality Effects in Dual Task Performance

Christopher D. Wickens, Joy Mountford, Diane Sandry and William Schreiner
University of Illinois Department of Psychology
Champaign, Illinois 61820

Abstract

Two experiments are reported which assess whether individuals differ in terms of a general time-sharing ability, and whether time-sharing performance differs as a function of the relation between hand assignment and processing requirements (spatial vs verbal) of the two tasks. The answer to the first question was generally negative. Specific, but not general, time-sharing skills were identified. The data concerning the second issue were equivocal, supporting the existence of laterality effects in one experiment, but not in a second one.

The efficiency with which two tasks can be time-shared varies as a function of both the individual performer and the specific tasks involved. Anecdotal evidence of individual differences in time-sharing may be provided by variations in the degree to which the skilled pilot, in contrast to the amateur, can interweave and time-share complex information processing tasks. Experimental evidence along these lines has been provided by prediction studies in which unique variance associated with laboratory performance in dual task conditions can serve as a useful predictor of flight training success (e.g., Gopher & Kahneman, 1971; Damos, 1978).

There are however two alternative interpretations to these findings. Postulation of a unique time-sharing ability suggests that individuals differ in terms of the total amount of processing capacity, or the flexibility and speed with which attention can be switched. These skills will only be manifest in a dual task environment, and so will not be tapped by single task measures. According to an automation hypothesis however, certain individuals can perform tasks with a smaller demand for limited resources (automation), but not necessarily at a higher level of proficiency. Such individuals therefore will not necessarily excel others in single task performance. Only when the limited resource supply is demanded by concurrent activity (in both the dual task testing conditions and in tests of flight proficiency) will the effect of this difference in automation be manifest. For the subject whose performance is more automated, these resources can be diverted with little cost to performance. For the non-automated group on the other hand, performance will be sacrificed; therefore performance differences will emerge (and co-vary) in the two conditions (dual task laboratory and inflight).

The point to be emphasized here is that this apparent individual difference in time-sharing ability is not really differences in attentional capacity or flexibility, but are specific properties of the individual task components. In order to tease apart these two sources of difference in dual task efficiency, it is necessary either to focus upon a microscopic analysis of differences in processing that might correspond to phenomena of switching speed or flexibility (Damos & Wickens, 1980; Keele, Neill & DeLemos, 1980) or alternatively, to determine if individual differences in time-sharing can be discovered that transcend a wide

variety of different dual task combinations. While previous investigations that have sought such a "transituational" time-sharing ability have not met with success (e.g., Sverko, 1976; Jennings & Chiles, 1977; McQueen, 1917), careful scrutiny of these studies indicates that all include some methodological or theoretical shortcoming that might have precluded the emergence of the skill (see Wickens, Mountford & Schreiner, 1979).[1]

Experiment 1: Individual Differences in Dual Task Performance

The purpose of Experiment 1 was to investigate the existence of a general time-sharing ability within an information processing framework. A factor analytic approach was employed in which 40 male subjects performed each of four information processing tasks singly, and time-shared with each other in a variety of pairwise conditions. The tasks consisted of (1) manual tracking; (2) a visual digit classification task; (3) a task requiring a spatial judgment of line orientation (estimating the imaginary intersection point of two converging lines); and (4) an auditory running short term memory task. The auditory task required a yes/no decision concerning whether two sequential letter stimuli were in correct sequential alphabetical order. The latter three tasks required a 1 bit key press response while tracking required continuous lateral manipulation of a control stick. On alternate days, subjects performed the tasks, both singly and simultaneously, with opposite hand assignments.

1. Individual differences in dual task efficiency. Two separate principal component analyses (PCAs) were performed on the data. One employed all performance scores (both single and dual) as variables, and the second was performed on the dual task efficiency scores from each time-sharing condition. In this measure the decrement of each task in a pair from its single task control is computed. The measures for a pair of tasks are either treated as a bivariate observation (see 2. below) or, in the second PCA are averaged. Since the performance measures are different for tasks within the pair, these decrements are converted to normal deviate scores prior to the averaging, by dividing by an estimate of the trial-to-trial variation.

The results from both PCAs generally substantiated the conclusions drawn by previous investigators concerning the existence of a general time-sharing ability. The primary factor loadings on the first analysis were restricted to tasks under both single and dual task conditions (no factor emerged that loaded only on dual task performance). Furthermore, in the decrement PCA there was no evidence for a "general factor" that loaded on a number of qualitatively different time-sharing conditions. Instead, the two factors that did emerge from this analysis related respectively to time-sharing of two visual tasks demanding of high visual acuity (presumably related to scanning strategies) and to time-sharing of the auditory task, paired with any other task. This factor was presumably related to the automation, or resource demands of the auditory task.

2. Task related differences in time-sharing efficiency. These differences were evaluated by performing a bivariate analysis of variance

---

[1] This research was sponsored by a grant from the Navy Aeromedical Research Lab. Detachment, New Orleans, La. Commander Robert Kennedy was the technical monitor.

on the dual task efficiency scores to assess differences in time-sharing efficiency of each task, when paired with itself,[1] and with each of the other three. Thus a separate MANOVA was performed for each task.

Of greatest interest in these analyses was the observation that task interference was a direct function of the number of processing structures shared between tasks. Tasks that shared common input modalities (auditory vs visual) or response types (discrete vs continuous) interfered with each other to a greater extent than tasks which did not. These effects are quite consistent with a multiple-capacity conception of dual task interference (Wickens, 1980). However, the dimensions that defined differences in time-sharing did not, as noted, define individual differences.

3. Laterality effects. As described above, the experiment was designed in such a way that task pairs were performed on successive days with opposite hand assignments. The importance of this design feature is inherent in the possibility that tasks may be shared more efficiently if the cerebral hemisphere which is processing information for a given task is the same one that is controlling the response. That is, a certain amount of "task integrality" is preserved among the processing and response requirements of each task. This integrality would be observed for a given task pair in which the processing of the two tasks is strongly lateralized. For example the digit categorization task might be expected, on the basis of hemispheric laterality data, to invoke processing of the left cerebral hemisphere, while tracking, because of its spatial processing requirements could be right hemispherically lateralized.

In fact, just such a laterality effect was observed in the tracking and digit task combination. Dual task efficiency on both tasks was significantly (and considerably) greater when tracking was performed with the left hand and the digits responded with the right, than when the opposite assignment was employed. This pattern of results thereby supports the hemispheric association of the two tasks and the importance of task integrality.[2]

Experiment 2: The Laterality Effect Pursued

Because of the potential system design implications of the laterality result with respect to the positioning of controls for various kinds of discrete and continuous tasks, it was pursued in a second, more controlled experiment. Subjects tracked (with either the left or right hand) singly and concurrently with a visually displayed Sternberg Memory Search Task. In this task, letters are presented, to be judged whether they are or are not a member of a previously memorized set. Like tracking, reaction times to the memory search task were also collected in both the dual task conditions and in a single task control with the handedness of response assignment manipulated in a complementary fashion. In this condition we expected to replicate the preceding results of improved time-sharing efficiency with a tracking-right, RT-left, hand assignment.

As an additional variable, a second form of the Sternberg Task

[1] Because of instrumentation difficulties, the auditory task was never paired with itself.

[2] This research was supported by a contract from the Office of Naval Research No. N000-14-79-C-0658. General Malecki was the technical monitor.

was employed, consisting of an "alphabet" of random dot patterns. The purpose of this manipulation was to further establish the spatial demands of tracking. If tracking were indeed spatial in its processing requirements, then greater interference should be observed when it was time shared with the spatially defined sternberg Task, than in the letter version of the task, by virtue of the common demands upon spatial processing resources (Wickens, 1980).

With regard to the spatial-verbal processing difference, our predictions were not confirmed, for the data on day 2, after performance had stabilized, time-sharing efficiency was no greater with the verbal than the spatial version of the Sternberg Task. However, the second investigation replicated the hemispheric-task integrality result observed in Experiment 1: for the verbal condition, in which integrity was provided with the tracking left-Sternberg right assignments, time-sharing efficiency (the difference between single and dual task performance on both tasks) was significantly better ($p < .02$) than with the non-integrity assignment.

The failure to observe a main spatial-verbal effect may have resulted from the adoption of verbal coding strategy by some subjects in the spatial task. Indeed Cohen (1979) has argued that laterality effects are sensitive to processing strategies involved. While subjects were requested not to name the spatial stimuli, it is possible that some might have done so, particularly in light of the fact that time-sharing efficiency could be improved as a consequence.

In summary, the results of Experiment 1, in support of prior investigations, appear to cast doubt upon the possibility that a general time-sharing ability does exist, or if it does, that it accounts for sufficient experimental variance to render it worth pursuing. The combined results of Experiment 1 & 2 suggest that the phenomenon of task-hemispheric integrity is probably a reliable one that has important potential implications for the configuration and placement of controls in system design.

References

Cohen, G. Comment on "Information processing in cerebral hemispheres". Journal of Experimental Psychology: General, 1979, 108, 309-315.

Damos, D. L. Residual attention as a predictor of pilot performance. Human Factors, 1978, 20, 435-440.

Damos, D. L., & Wickens, C. D. The identification and transfer of time-sharing skills. Acta Psychologica, 1980 (in press).

Gopher, D., & Kahneman, D. Individual differences in attention and the prediction of flight criteria. Perceptual & Motor Skills, 1971, 33, 113-115.

Jennings, A.E., & Chiles, W.D. An investigation of time-sharing ability as a factor in complex performance. Human Factors, 1977, 535-548.

Keele, S. W., Neill, W.T., & DeLemos, S.M. Individual differences in attentional flexibility. University of Oregon Center for Cognitive and Perceptual Research. Tech Report, 1979.

McQueen, E.N. The distribution of attention. British Journal of Psychology, 1917, II. (Monograph Supplements 5).

Sverko, B. Individual differences in time-sharing performance. University of Illinois Aviation Research Lab. Tech. Report No. ARL-77-4/AFOSR-77-4, 1977.

Wickens, C.D. The structure of attentional resources. In R. Nickerson (Ed.) Attention and Performance VIII, Erlbaum, 1980.
Wickens, C.D., Mountford, S.J., & Schreiner, W. Task dependent differences and individual differences in dual task performance. University of Illinois Engineering Psychology Laboratory Technical Report EPL-79-4/NAMRC-79-1, 1979.

ESTABLISHING AND MAINTAINING AN OUTREACH PROGRAM IN BASIC TRAINING

Thomas P. O'Hearn
Wilford Hall USAF Medical Center
San Antonio, Texas 78236

Abstract

In October 1976 the Mental Hygiene Clinic, of the Department of Mental Health, Wilford Hall USAF Medical Center began a community outreach program in the Basic Military Training School Squadrons. Mental health technicians were assigned to cover one or more basic training squadrons for the administration of the Air Force Medical Evaluation Test program. In addition, they were to serve as mental health consultants to squadron personnel. As such they were involved in early assessments of basic trainees who for various reasons were having difficulty functioning in the basic training environment.

The management difficulties encountered in the establishment and maintaining an effective outreach program centered around problems of communication with the training squadrons, manpower fluctuations in the Mental Hygiene Clinic, and high turnover of trained mental health technicians. The latter two problems have affected the ability of the clinic to provide the established mental health services as well as to begin innovative interventions.

For the past two decades mental health practitioners have been focusing on interventions in the community. The shifting of attention from the clinic to the community has brought about changes in the delivery as well as the type of mental health services offered. The reader is directed to books by Cowen, Gardner and Zax (1967), Iscoe and Speilberger (1970), and Korchin (1976) for a more complete picture of the history, development and current status of the field. This paper will briefly address the development of an outreach program at the Basic Military Training School, Lackland Air Force Base, San Antonio, Texas. The major issues to be addressed in both the establishment and maintenance phases of the program are: manpower; training; types and levels of interventions.

Establishment Phase

Prior to October 1976, the Mental Hygiene Clinic functioned, for the most part, as a traditional outpatient mental health clinic. Basic trainees who, for a variety of reasons, were having difficulty adjusting to basic training were referred to the clinic primarily by squadron personnel. Often, by the time the trainee was seen by Mental Hygiene, he or she was beyond assistance. Thus, for many basic trainees, the Mental Hygiene Clinic was one of the last stops in the process of being discharged from the Air Force. Those who were not recommended for discharge were usually returned to the squadron with little or no provisions for followup. A small percentage of trainees were hospitalized following assessment by the clinic staff. Hospitalization usually ended in the trainee being eliminated from the service (McCabe and Board 1974).

PRECEDING PAGE BLANK-NOT FILMED

The manning of the clinic during the past decade changed in complexion from a largely officer staff (four psychiatrists, one psychologist, and four social workers, several psychiatric technicians, and two secretaries) in the late sixties to a largely technician staff (one psychologist, two social workers, one psychiatric nurse, six psychiatric technicians and two secretaries) of the early seventies. This model persisted for the remainder of the decade with the psychiatric nurse being replaced by a Master's level psychologist. The population served approximately six thousand basic trainees and fifteen hundred tech school students per month. The clinic also serves the seven hundred training instructors and their families.

In 1972, several department of Mental Health staff of Wilford Hall USAF Medical Center developed a screening test, the History Opinion Inventory (HOI). The purpose of the test was to identify, early in the training process those basic trainees who would be a risk for effective service (LaChar et. al., 1974). During 1973 the test was validated on 14,000 basic trainees and found to be fairly predictive of poor performance in the Air Force (Guinn 1975). In June 1975, the Air Force Medical Evaluation Test (AFMET) program was begun as a three phase assessment research project incorporating the HOI (Phase I), a structured interview and the MMPI (Phase II), and a formal mental hygiene evaluation (Phase III). The results of the year long research program were encouraging. The AFMET program was able to identify airmen who were ultimately recommended for discharge before completion of basic training (Bloom 1976).

The Air Force decided to make the program operational in October 1976. The operational phase of the program was attached to the Mental Hygiene Clinic with the addition of nine technician positions for the AFMET program. This addition made possible the placing of mental health technicians in the basic training squadrons. In the squadron setting they administered the Phase I (HOI) and Phase II of the AFMET program, (the Bloom Sentence Completion Test, which replaced the MMPI, and a structured interview.) In addition, they would also provide other mental health services to their assigned squadrons (O'Hearn, 1978).

Placing the technicians into the squadron was a gradual undertaking. This was due, in part, to the clinic's gradual manning buildup. Also, while instituting this new service to the basic training community, it was necessary not to disrupt the services to which the community had become accustomed. Two technicians, one senior staff sergeant with ten years mental health experience and one senior airman who was working on a Master's degree in social work, were selected to be integrated into two of the twelve basic training squadrons. The two squadrons were selected because the technicians had already established a degree of rapport with the squadron personnel. It was seen as important that the initial placement within the squadrons be a positive experience for both the technicians and the squadrons.

In addition to providing timely mental health services to these two squadrons, the two technicians provided excellent training for other technicians who followed them into the basic training squadrons. The placement of the follow-on technicians in particular squadrons was done, when possible, with the cooperation and agreement of both the technician

and squadron. Also, technicians were not put into the squadron setting until both the technician and supervisor were comfortable with the placement. This was very important for technicians new to the career field.

The movement of the technicians into the squadrons presented some problems with regard to supervision and identity. To deal with the problems of supervision four teams were formed. Each team had an officer and senior NCO in supervisor/consultant roles to several squadron technicians. In reality, such consultation could be obtained from any of the officers and senior NCO staff by any of the squadron technicians.

The problem of technician identity presented different problems in different squadrons. Most squadrons looked upon the technician as a consultant who did not "belong" to the squadron. This viewpoint was one with which the clinic was most comfortable. However, there were some squadrons in which the technician was subtly pressured to become a squadron member. This was not seen as deliberate on the squadron's part, but merely their way of being most comfortable with personnel in their environment. Squadrons also become protective of their technicians and, at times, voiced opposition to moving a technician with whom they had worked out a good relationship. It became very important to closely coordinate with the squadrons both the placement and the removal of squadron technicians.

The establishment phase of the outreach program took approximately one year to accomplish. Problems in maintaining this outreach program will be the subject of the remainder of the paper.

Maintenance Phase

There are two main underpinnings of the Mental Hygiene Outreach Program. One is the AFMET program which provides the manpower authorizations for the clinic as well as a large patient count for the department. The second is the commitment by the medical center, Mental Health Department, and the Mental Hygiene Clinic to provide quality mental health services to the basic training community. The clinic's philosophy is that we are in the basic training community to help the training personnel do their jobs better. Clinic staff must keep in mind that if Mental Hygiene should close, BMT would still graduate trainees -- hopefully, not as efficiently. Basic training would certainly not close down.

It is important for both clinic and serviced community to keep mindful of the nature of the relationship between the Mental Hygiene Clinic and the basic training school. There needs to be a mutual respect for the other's integrity as well as missions.

Maintaining an effective outreach program necessitates communication on several different levels. Given the nature of the environment, military rank plays an important role in the communication processes. It appears that within both systems, upward and downward communication patterns are well established. Between the two systems there is a great deal of peer interaction -- officers with officers, NCO's with NCO's, etc. However, there is also a fair amount of communication across rank lines between MHC and BMTS personnel, especially in regard to diagnosis, treatment, and recommendations as to disposition of basic trainees. As with any organization, communication between clinic members is of great importance. Due to the diversity of the responsibilities of the squadron technicians, coordination of their activities with clinic and squadron personnel is essential.

Breakdown of coordination and communication has proven to be one of the major problem areas in managing the outreach program. This is especially true when one technician must cover more than one squadron which, because of manning shortages, has been the norm.

Maintenance of an outreach program also depends on the ability to maintain some stability of the staff. The lack of staffing stability had been, from a management point, a most frustrating facet of the Lackland program. During the three operational years there has been a three hundred percent change in the manpower picture. This has put a strain on the ability to provide continuous, efficient, and effective services to basic trainees and squadron personnel. This lack of ability to control the stability of personnel makes planning and organizing difficult. The management process seems, at times, an exercise in futility. An example of this occurred during the Summer of 1979. Three technicians were transferred to Mental Hygiene from other areas of the Mental Health Department to help fill six clinic vacancies. Within three months of being assigned to the clinic, two of the technicians had orders for Germany. In addition another staff member was transferred overseas. Efforts to salvage at least one of the positions met with failure.

The lack of control of manning resources also leads to an attitude of holding back on initiating worthwhile programs or honoring reasonable requests by the community. With limited resources, effective maintenance of an outreach program does necessitate sometimes saying "no" to requests for services by community agencies. In the words of the comedian Flip Wilson, "Don't write a check with your mouth that your body can't cash." In the case of the Mental Hygiene Clinic at Lackland, this meant providing only evaluative services to the tech schools, severely limiting the number of therapy contracts entered into by staff, referring dependents to the outpatient psychiatric clinic, and dropping non-essential clinical services to trainees such as a bi-weekly group when the clinic manning could not realistically support these activities.

On-going training of clinic personnel is seen as important in maintaining an effective outreach program. The squadron technician in the outreach program are asked to do what might be done by a master's level individual in civilian life. There is a need to constantly upgrade the technicians through formal courses, discussions, lectures and supervision, as well as informal day to day contacts. The outreach program at Lackland is fortunate in that the department of mental health has residencies in psychiatry and psychology which provide abundant training opportunities. There is also an active distinguished visiting professor program which brings in national leaders in mental health to Wilford Hall. Clinic personnel can and do take advantage of the numerous training opportunities from these department programs.

Another important part of the outreach program which helps in maintaining effectiveness is an operational research program. The clinic has such a program, AFMET, which has provided excellent data from a management point of view. With limited resources, it is important to know that interventions and programs are most effective with what target populations. Without feedback, good programs can dilute their effectiveness by spreading services too thin or directing services to populations who will not significantly benefit by interventions.

Types and Levels of Intervention

The two main types of intervention carried out by the clinic are those generated by the AFMET program and those generated by several base agencies. Due to the fact that most of our patient population is on base for only about two months, early identification of problems and quick interventions are necessary. Crisis intervention and short term therapeutic techniques are utilized almost exclusively. There are a few opportunities for longer term interventions when the basic trainee will be assigned to Lackland for tech school or as a permanent assignment. For individuals who may need followup mental health services after leaving Lackland, arrangements are made with mental health personnel at the gaining base to provide these services. This insures a quick entre into the new base's mental health system.

Interventions with permanent party personnel are usually more flexible as to modality used. However, as noted earlier, the ability of the clinic to carry a large, long-term patient population is limited by the need for quick response to emergent problems in the highly transient trainee population. Therefore, most of the officer staff carry only a few long term cases. Requests for long term individual treatment,marital or family therapy usually are referred to the outpatient psychiatric clinic at Wilford Hall.

Besides providing direct services to basic and tech school trainees, the clinic personnel also consult with those who come in direct contact with the trainees. The clinic has had formal input to the instructor training program and both formal and informal consultation is available to training instructors in the squadron areas. The clinic director regularly attends the staff meeting of the basic training squadron commanders and consults with squadron commanders and school command staff on both individual and systems oriented problems. The total impact of having this interface by clinic personnel has resulted in what is perceived by both MHC and BMTS as significant changes in the basic training environment. There is more of a team approach to solving people and systems problems with continued cooperation as a goal which will benefit the training personnel, trainees, and the clinic. An example of this cooperative effort is the emphasis by both the clinic and school in early identification of those trainees who are risk for acting out in a self destructive manner. During the period that mental health technicians have been in the squadrons there have been no suicides in the basic training community. In addition, the rate of suicidal incidents, almost all of which were of low lethality, went from approximately one hundred and fifty per year prior to 1976 to seventy-one per year for the past two years.

In summary, the placing of mental health technicians in the basic training squadrons beginning in October 1976 has enabled early intervention in crisis and potential crisis situations. The unique problems faced by the training squadrons also presented some unique problems for the delivery of mental health services. It took about a year for the outreach operation to run smoothly. Recurrent problems revolving around shortages in clinic manpower and the seeming constant rotation of clinic and training school personnel were major irritants and blocks to effective service. Continued communication and cooperation between clinic staff members, the clinic and the serviced community are the behaviors that make for a viable and effective outreach program.

References

Bloom, W. Air Force Medical Evaluation Test, Medical Service Digest, 1977, XXVIII, No. 2.

Cowen, E. T., Gardner, E. A., and Zax M. (Eds.) Emergent approaches to mental health problems. New York: Appleton-Century-Crofts, 1967.

Guinn, N., Johnson, A. C. and Kantor, J. E. Screening for adaptability to military service. Report No. AFHRL TR-75-30, Air Force Resources Laboratory, May 1975.

Iscoe, I. and Spielberger, C. D. (Eds.) Community psychology: perspectives in training and research. New York: Appleton-Century-Crofts, 1970.

Korchin, S. K. Modern clinical psychology principles of intervention in the clinic and community. New York: Basic Books, 1976.

LaChar, D. Sparks, J. C., Larsen, R. M. & Bisbee, C. T. Psychometric prediction of adaptation for USAF basic trainees: Journal of Community Psychology, 1974, 2(3), 268-277.

McCabe, M. & Board, G. Stress and mental disorders in basic training; Military Medicine, October 1976. pp 686-688.

O'Hearn, T. The Air Force Medical Evaluation Test (AFMET) Program: A three phase assessment program. Proceedings: Psychology in the Department of Defense, April 1978, pp 261-264.

The Human Development Center, Philosophy and Practice

James L. Jenkins
Director, Human Development Center
Lowry Air Force Base, Colorado 80230

Abstract

Limitations within the mental health care delivery system resulted in significant gaps in meeting professional service requirements among technical training students at a substantial cost to the USAF. Human Development Centers were implemented at three technical training centers to provide preventive mental health services. Systems theory and the preventive health care continuum provided a conceptual framework for the development of a full range of prevention programs implemented at all system levels within the training environment.

In response to significant gaps existing in services provided within the mental health care delivery system, Human Development Centers (HDC) were implemented at three Technical Training Centers during 1979. This paper presents a background on conditions and events that served as a catalyst for the development of the HDC, summarizes Air Training Command (ATC) actions that gained United States Air Force (USAF) sanction and resources, identifies the conceptual framework underlying program development, and discusses planning strategies that were utilized for program implementation.

Background

Two variables that existed within the technical training environment that set the stage for the HDC were first, the nature of the patient population and their associated mental helath service requirements and second, the limitations of existing mental health care delivery systems.

The Patient Population

The characteristics and mental health requirements of the technical training students, hereafter referred to as students, were significant variables in the identification of the need for the development and implementation of the HDC.

Characteristics. In general, students are 18 to 22 years of age, are high school graduates with college or work experience who joined the USAF to obtain technical skills and educational opportunities. They were successful during the pre-enlistment years and are motivated to succeed in their military enlistment as a means for future self-actualization.

Mental health care requirements. A small percentage of students present with significant psychopathology. Most cognitive, affective or behavioral disorders among this population are transient disturbances of a developmental or situational nature associated with such factors as separation, military structure and control, academic demands or failure, and social or interpersonal processes. Manifestations range from anxiety and depression to psychophysiologic disorders and maladaptive behavior including substance misuse and suicidal gesturing.

The Mental Health Care Delivery System

Mental health care delivery systems were identifed as having four limitations that contributed to their failure to meet the mental health care requirements of the student population.

Entree. Entree into the mental health care delivery system is perceived as closed and unaccessible by most individuals requiring professional assistance. Resistance is kindled and access blocked by misunderstanding, fear, public opinion and a network of appointment and screening systems. Contact is often delayed until an appointment can be arranged. Consequently, minor transient conditions remain unattended until they have escalated to a point of acute and severe dysfunction.

Treatment philosophy. Many mental health clinicians have a psychoanalytic treatment philosophy with a personal interest in providing therapy to individuals manifesting signficant psychopathology. At technical training centers, mental health clinicians are required to spend a high percentage of their time providing diagnostic and dispositional decisions in response to transient disturbances among students. Consequently, clinicians become frustrated and, in turn, increasingly insensitive and unresponsive to the needs of the students.

Prevention effort. A number of variables beyond the control of individual mental health clinicians have contributed to a systematic exclusion of any meaningful primary prevention effort within the mental health care delivery system. The primary variable is the "patient count work load factor" which determines staff authorizations based on the number of patient contacts. This procedure penalizes any meaningful effort toward primary prevention and perpetuates the focus of mental health services on the individual patient.

Degree of mission support. With mental health resources being distributed among active duty and retired military members and their families, a limited amount of these resources are devoted directly to the support of the mission of the installation to which the medical facility is attached. Consequently, students compete with all other segments of the total patient populations for services.

Cost

The limitations of the mental health care delivery system have resulted in a significant cost to the USAF. Some of the costs associated with these limitations were identified and quantified for 291 students receiving mental health services at Lowry AFB during 1978.

Delayed identification. The failure to identify conditions among students at the time of onset results in a period of limited productivity while the student remains in training. The mean lag time from onset to identification was 28 days resulting in 8,148 training days with limited productivity and signaling a need for early identification of dysfunction among students.

Delayed intervention. Once a student is identified as requiring mental health services they are generally removed from training pending evaluation. The mean lag time from identification to mental health intervention was five training days. This delay resulted in 1,455 training days at $31.51 per day at a total cost of $45,847.

Delayed administrative responses. Of the 291 students receiving mental health service during 1978, 135 required written response to the mili-

tary unit. The mean lag time from completion of the evaluation to the unit's receipt of the report was 13 days resulting in a total of 1,755 ineffective training days at a cost of $55,300.

Unnecessary physical health care. Many students experiencing psychological or emotional distress present at "sick call" due to psychophysiological reactions or a desire to talk with someone who will listen and understand. Many of these visits could be prevented with a responsive mental health care system. During 1978, 71 students presented for 110 outpatient visits at $29.00 per visit at a cost of $3,190.

Inappropriate requests for mental health services. During 1978, 69 of the requests for mental health services by military units were considered inappropriate in that they were associated with problematic behavior that should have been managed within the unit. These requests generated a requirement for unnecessary services at a cost of $2,001 and demonstrated a need for consultation services to unit commanders on the management of students.

Attrition. Delayed intervention often results in a requirement to discharge students who might have been retained with more responsive mental health services. During 1978, 94 students were recommended for separation by mental health, 50 of which might have otherwise been retained. The USAF had invested an estimated $150,000 in training these students prior to their discharges.

Summary. The combined cost of these factors at Lowry AFB during 1978 is estimated to be $256,257. Assuming that a similar situation existed at the other three technical training centers, the total cost of limited mental health services to the USAF during 1978 was approximately $1,025,128. In addition, there were unquantifiable excessive psychological and emotional costs experienced by these students as a result of these limitations.

Requirement for Primary Prevention Established

During 1976 the ATC Surgeon, Brigadier General Joseph E. Wesp, conducted a study to assess the extent of primary health care throughout the command. The study findings identified the need to provide primary mental health care for the student population. A proposal for the establishment of HDCs was developed and an aggressive campaign initiated to secure the required resources for implementation. As a result of this effort, the USAF authorized 20 staff positions for the implementation of a primary mental health prevention program on a pilot basis.

Guidelines. In November 1977 General Wesp stated that he and his staff were dedicated to the establishment of Human Development Centers and directed that they be developed within the following guidelines:

1. To provide essential preventive mental health care.
2. To assure rapid identification and resolution of student problems.
3. To provide required adjustment services to students.
4. To provide consultation to command and training staff.
5. To identify stress vectors which hinder learning.
6. To reduce attrition from training and the USAF.

Objectives. Subsequently, key people involved in planning the HDC defined the following program objectives:

1. To ensure the emotional stability of students and staff.
2. To maximize the training experience for all students.

3. To minimize student attrition.

Conceptual Framework

The conceptual framework utilized in the development of a primary prevention model for the HDC is based on a preventive matrix which incorporates systems theory and the preventive health care continuum.

Systems Theory

The systems concepts that are relevant to the development of the HDC are first, that all components within a system are part of an interdependent network of supra-systems and sub-systems and second, that each component influences the others with a change in any part of the system having an impact upon the total system. The Technical Training Wing (TCHTNGWG) is conceptualized with four sub-systems or system levels.

Student system. Each individual student arrives for training with a different degree of social maturity, psychological development and academic background. HDC provides services to the student system to enable each student to complete training as a job competent technician.

Training system. The training system is composed of military and academic personnel who are directly involved in providing students with military training and academic instruction.

Command system. The command system consists of command and staff personnel who participate in decision making, policy determination and program development that has a direct bearing upon the training process.

Support system. The support system is composed of agencies or units responsible for providing professional and support services to the TCHTNGWG.

Prevention Continuum

Providing services within the full range of the prevention continuum is another important concept in the development of the HDC.

Primary prevention. Primary prevention services are provided to all components and system levels within the TCHTNGWG to maintain the efficiency and effectiveness of the training mission.

Secondary prevention. Secondary prevention services are provided to components of the TCHTNGWG that are identified as "high risk" in order to prevent the occurrence of probable dysfunction.

Tertiary prevention. Tertiary prevention services are provided to components where dysfunction has occured in an effort to resolve the dysfunction and prevent a reoccurrence.

Preventive Matrix

The conceptual framework for the HDC consists of an integration of the four system levels and the prevention continuum. HDC programs provide primary, secondary and tertiary prevention services to the student system, the training system, the command system and the support system.

Implementation Strategies

The strategies that were utilized during the planning and development phases of the HDC are considered critical for the implementation and survivability of the program.

Manpower Standards

The "patient count" work load factor utilized for staff authorizations within the medical corps prohibits preventive health care. Consequently, the implementation and survival of the HDC requires the utilization of a

different work load factor. ATC management engineering personnel conducted a manpower standards study of the work center description for the HDC and recommended utilization of a work load factor based on the number of students in training. Air Force acceptance of this recommendation is considered essential for the future survivability of the HDC.

Psychological Ownership

The successful implementation of the HDC required psychological ownership of the program by key people within the TCHTNGWG. This was achieved by soliciting and maintaing the involvement of these key people throughout the planning, development and implementation phases of the program. As a result, the sanction, resources and support that were critical for successful implementation were provided.

Interface Within the Training Environment

A final strategy was the relocation of HDC personnel from the medical to the training environment. HDC offices were established in the training area where staff members interface with students and training staff in the dormitories and classrooms.

Program Development

The development of HDC programs with a focus on primary prevention required a radical shift from traditional diagnostic and treatment services to educational, consultation and personal adjustment services.

Preventive Education

A significant component of the primary prevention effort of the HDC is educating students and staff about psychological and emotional processes and the management of their personal responses to stress associated with technical training. This educational process is accomplished through briefings, workshops and seminars that are conducted for both students and the training staff.

Consultation

A major thrust of the HDC program is the provision of consultation services to training and command staff. Consultation is designed to assist them in the identification and management of student problems as well as understanding the emotional impact of training decisions and policies.

Adjustment Services

In addition to the prevention of developmental and situational adjustment reactions among the student population, the HDC is also designed to assure the early identification and rapid resolution of such conditions when they do occur. Therefore, professional services are provided to assist students in developing coping skills through participation in individual and group adjustment programs. The focus of adjustment services is on dysfunction associated with the technical training process.

Conclusion

The implementation of the HDC has marked a significant advancement in preventive health care within the USAF. The physical location of HDC personnel within the training environment resulted in increased visibility and accessibility. Preventive services are provided at all system levels in support of the training mission. Continuation of the HDC will require implementation of the recommended work load factor and the allocation of required resources in the future.

The Human Development Center Within the Department of Mental Health - The Sheppard Model

John H. Stokes
Director, Human Development Center
Sheppard Air Force Base, Texas 76311

Abstract

Opened in November 1978, the Sheppard Human Development Center (HDC) was organizationally structured as a subunit of the Department of Mental Health. It is the only HDC providing comprehensive mental health services to its catchment population. While it provides all traditional outpatient and inpatient services, its primary focus is upon preventive programs that are based on the philosophy that people are healthy and dynamic entities who have a need to grow, learn, and develop to their full capacities. Its programs are designed to reflect this philosophy and to promote efficient, effective, productive and job satisfied persons fully committed to the Air Force. Examples of the services provided under its major programs are given.

The Sheppard Human Development Center (HDC) was formally opened in November 1978. It is organizationally structured as a vertical subunit of the USAF Regional Hospital's Mental Health Department, but is physically separated and located in an area of high student traffic and activity.

The HDC is staffed by four master's and one doctoral degree clinical social workers, six mental health technicians, and two civilian secretaries. In addition, there are two half-time clinical psychologists and two psychiatrists. The staff is fully dedicated to the HDC philosophy, purposes, and military mission. The staff adapt a dynamic behavioristic and humanistic philosophy in which the person and other systems are viewed as ever-developing, ever-learning, ever-growing, and ever-changing entities ultimately responsible for their own health, behavior, actions, attitudes, and growth.

The didactic, experiential, and therapeutic programs implemented and offered at the HDC are designed so that their boundaries interface with and are integrated with one another in such a manner as to (a) be preventive, educative, therapeutic, growth provoking, and cost effective; (b) minimize environmental, interpersonal, and intrapersonal stresses; (c) provide an ongoing assessment of the social, psychological, cultural, economic, vocational, and organizational factors affecting the individual, the family, the group, and the organization; (d) mobilize and utilize all available environmental support systems to favorably influence the conditions under which members live, learn, work, and interact; (e) favorably impact upon the organization at all levels by producing and/or enhancing an organizational climate conducive to the training and retention of effective, efficient, job satisfied, and productive members fully committed to the Air Force and its mission. The HDC's major programs are described in the following paragraphs.

Evaluation Program

The HDC staff provide full diagnostic services including assessments of the person's mental health and/or illness; personality; behavior, social,

PRECEDING PAGE BLANK-NOT FILMED

academic, and vocational skills and functioning; and a wide range of developmental needs and desires. The evaluation includes a prognostic statement of the person's motivation and capacity to adjust and contribute to the Air Force, and a specific vocational and/or academic speciality.

Depending upon the results of the evaluation, the person may be admitted to a psychiatric inpatient ward, or recommended for (a) separation from the service; (b) retraining into another vocational speciality; (c) removal from technical school and placement in a direct duty assignment for on-the-job training; and/or (d) placed into one of HDC's programs. Such recommendations are humane for the individual in that they remove the person from or help the person cope with stress. They are preventive in that they help resolve disciplinary and/or emotional problems and their carry over affect on other military members. They are cost effective in that the member is transferred from a nonproductive to a productive role, or removed and replaced with a productive member.

Adjustment Services

The HDC staff recognize that, at times, a system's equilibrium may be disrupted due to real or perceived environmental stress; intrapersonal, interpersonal and/or organizational dysfunctioning; lack of social, physical, communications, vocational, intellectual, academic, and psychological skills; family problems; job dissatisfaction; homesickness; and a variety of other reasons. The disruption of one system's (e.g. the individual or micro system) equilibrium may disrupt that of another system's (e.g. the squadron or meta, and the macro or Air Training Command). This is especially true in the Air Force where system boundaries are so interfaced that the disruption of one system's equilibrium can negatively affect the equilibrium of all others.

Adjustment services are designed to be humane, preventive and cost effective in that they (a) enhance the member's self-esteem and avoid the stigma attached to failure and/or separation from the service; (b) reduce behavioral or disciplinary problems within the squadron; (c) allow training staff to spend more time rewarding and enhancing the growth potential and job satisfaction of non-problem students; (d) reduce stress and anxiety for the member and training staff; (e) provide the member with help in adjusting to the Air Force or alleviating the anxiety associated with separation from the service; and (f) reduce training costs by keeping the member in his/her assigned speciality and by allowing the Air Force to avoid expenditures of retraining the member or replacing him or her with a new member.

To help individuals adjust, the HDC offers: (a) individual counseling and psychotherapy; (b) four ongoing weekly adjustment groups; (c) two ongoing marital counseling groups; (d) marital counseling for individual couples; (e) family therapy; (f) a relaxation therapy group for persons experiencing job-related or other stress; and (g) a study skills adjustment group which is designed for students who need to learn new study skills, how to pay attention, how to concentrate, and how to adjust to the stress of technical school and/or the Air Force.

Squadron Liaison Program

As a meta system, the student training squadron and its staff interface with and serve a vital "linking pin" function between the micro, the macro, and the supramacro systems. The squadron environment and personnel

have a tremendous affect on the student's adjustment to technical school and the military. Consequently, it is essential that squadron personnel operate at a high level of physical, intellectual, social, vocational, and emotional competence. This requires that they understand organizational dynamics; role models; human behavior, personality, needs, motivation, and growth and development; job satisfaction and productivity, group dynamics; supervisory and counseling methods and techniques; organizational morale and climate; and a variety of other factors. To help meet such needs, the HDC staff devotes a major share of its time, energy, resources, knowledge, skills, programs, and services.

The HDC squadron liaison program combines all aspects of the total HDC program. It includes consultation, evaluations, research, individual and group adjustment services, and educative workshops and seminars. The focus of the liaison program is upon early identification and resolvement of existing or potential problems. The program thus is concerned with prevention, adjustment, and cost effectiveness.

One professional and one paraprofessional are assigned as liaison personnel between the HDC and each of the two or three squadrons which compose a Group. The team is available to help squadron and/or Group personnel identify, research, assess, and verify existing or potential problems as well as to be involved in the problem-solving process as it involves, affects, and is affected by individuals, families, groups, other organizations, and the organization itself. The squadron liaison program operates in several ways: (a) the professional provides weekly formalized consultation to the squadron commanders, student training advisers, and first sergeant; (b) the professional informally visits with various students in order to learn about their school and squadron environments, problems, social events, adjustment to the military and a variety of other factors; (c) the professional provides a weekly didactic and experiential training and stress reduction group for current and potential student ROPE leaders; (d) the professional sets aside approximately four hours each week during which time students can drop in or be referred by school or squadron personnel on an informal basis to discuss academic, career, military, interpersonal, intrapersonal, and social needs, concerns, and attitudes; and (e) the paraprofessional spends approximately four hours a week within each squadron during which time students can drop in or be referred for counseling. Should a student need more intensive counseling or therapy, he or she is placed in the appropriate program at the HDC. As a result of its observations and interactions with both staff and students, the liaison team is able to identify potential and/or existing problems which it feeds back to and makes recommendations to commanders for resolution.

Consultation

Consultation is viewed as a two-way interaction--a process of seeking, assessing, giving, and receiving help in mobilizing internal and external resources to obtain the maximum effectiveness, efficiency, job satisfaction, and commitment of organizational members. The HDC staff provides ongoing consultation to: (a) crisis intervention organizations such as the Command Post Help Line, the Rape Crisis Team, and the Crisis Pregnancy Hotline; (b) the Drug and Alcohol Abuse Rehabilitation Committees; (c) Human Services Advisory Council; (d) Student Affairs Committee; (e)

School of Health Care Sciences programs such as the Nursing Management and the Physician Assistant programs; (f) Social Actions; (g) Base Chaplains; (h) Group Commanders; (i) Federal Working Women's program; (j) the Child Advocacy Committee; (K) the Young Married Airman's Association; and (l) the Technical Instructors Course; and (m) instructors, supervisors, commanders, faculty development staff, and professional and laymen groups on an "as needed or desired" basis.

Educative and Growth Programs

The HDC staff adapt the philosophy that (a) awareness, knowledge, growth, and job satisfaction are essential elements of any preventive mental health approach in an industrial setting such as the Air Force; (b) a self-aware and "growing" person is more healthy, motivated, effective, well-adapted, job satisfied, and committed to the organization; (c) a mature, self-aware, and growing person has a more positive affect on and contribution to other systems; and (d) people are healthy and have a need to grow, to increase their skills, and to positively impact on other systems. The educative programs reflect this philosophy and consist of briefings, workshops, and growth groups.

The HDC staff provides ongoing and one-time only briefings to individuals, groups, and organizations to increase (a) awareness of the HDC programs and services; (b) awareness of problems indicating referral; and (c) early referrals. For example, on the third day after they arrive at Sheprard, all new students attend a one-hour briefing by an HDC staff member. This briefing is to increase their awareness of (a) adjustment problems likely to be encountered at Sheppard; (b) veneral disease; (c) rape, (d) pregnancy; (e) interpersonal relationships; (f) self-responsibility; and (g) the HDC's availability and desire to help them with their problems and needs. This time is also utilized to have a random sample of students fill out a problem rating form that is utilized to research problems encountered by students before and immediately upon their arrival at Sheppard. This briefing is held Monday through Friday and is normally attended by approximately 100 students and results in an average of five self-referrals to HDC.

Briefings to training and squadron personnel result in (a) increased communications between the HDC and other military organizations and members; (b) a request for services such as relaxation, self-awareness, and assertiveness groups, supervisory and counseling skills workshops, (c) an increased awareness of the need for early referral; and (d) an increase in referrals. Early identification is important since the earlier the referral, the more successful the service, and the more preventive of further problems for the individual and other systems.

The HDC staff provide (a) a weekly four-hour supervisory and counseling skills workshop that focuses upon the counseling and supervisory problem-solving process, elements of effective supervision, subordinate needs and job motivation, conditions conducive to job satisfaction and performance, counseling techniques and skills, and behavioral and emotional warning signals warranting referral; and (b) a weekly four-hour character and behavior workshop that focuses upon ways to recognize normal versus abnormal behavior, attitudes, and actions.

The first workshop above is also conducted for the Advanced Pharmacy Leadership course. The latter workshop is also conducted for the Technical

Instructors and the Nursing Management courses.

In addition to briefings and workshops, the HDC staff conduct didactic and experiential growth groups designed to develop (a) self-awareness and responsibility; (b) recognition of and techniques to alleviate stress and job dissatisfaction; (c) and enhance communication skills; (d) sensitivity to the needs of others; (e) enhance interactional, counseling, and supervisory skills; and (f) improved knowledge about particular subject (e.g. mental health, personality disorders, etc.). Utilizing the above designs the HDC staff provide (a) an ongoing group for physician assistants that is also designed to enhance their knowledge of and treatment skills for mental illnesses; (b) a relaxation skills acquisition group for squadron and training staff personnel that is designed to teach relaxation and stress reducing skills utilizing progressive relaxation, desensitization, and biofeedback techniques and equipment; (c) two marital enrichment groups each of which is designed for marital couples who desire to expend time and energy towards more fully understanding themselves and each other; (d) two groups for individuals each of which is designed to enhance their self-awareness; and (e) a group for Medical Services Specialist students and a group for Mental Health Ward Technicians students each of which is specifically designed to reduce the high job dissatisfaction and attrition rates formerly encountered in these courses.

When the demand is present, the HDC staff also provide other programs and services (e.g. parental enrichment workshops) and expand the above briefings, groups, and workshops.

Research

The HDC staff has a viable and ongoing research program to: (a) assess the effectiveness of its own programs, services, and personnel; (b) identify potential and existing problems; (c) assess members' job satisfaction and morale; (d) assess members' needs and desires for new programs or changes in existing ones; and (e) actively participate in macro research projects with all other HDC's.

Existing instruments are utilized to assess the effectiveness of growth and adjustment groups, job satisfaction, morale, etc. New instruments are also designed, constructed, tested, and utilized to measure the effectiveness of specific workshops, consultations, and squadron liaison programs.

The Initial Operation of a Community - Primary Prevention Program at Keesler Air Force Base

Captain Henry J. Lagarde, Ph.D.
Clinical Psychologist
Human Development Center
Keesler AFB, MS 39534

Abstract

In July, 1978 the initial stages of a community-primary prevention program were instituted at Keesler AFB. This program became known as the Human Development Center (HDC), its objective being the support and enhancement of Keesler's training mission. A systemic and personal growth philosophy culminated in a community program organized into four phases: (a) Orientation of the training community to HDC (b) Professional interventions geared for social system levels (c) Identification of problems and trends in the training community, and (d) Process review of HDC impact on the training community. Principle modalities for the implementation of these phases are consultation, education, and experiential learning of social system skills. Professional attention to student adjustment difficulties continues as an important HDC responsibility.

In July, 1978, Keesler AFB began to concentrate efforts for the establishment of a primary prevention program for continued enhancement of Keesler's training mission. Increased emphasis on primary prevention was fostered through emerging concepts appropriate to the sensitivity of social-ecological issues and the encouragement of personal growth. An epistemological perspective recognized at Keesler is General Systems Theory (von Bertalanffy, 1968). Systems theory understands the "self" as an inclusive part of social-ecological "suprasystems" (Glad, 1959; Milsum, 1972; Angyl, 1941.) By teaching individual sensitivity to structured interpersonal, group, and community relationship patterns, it becomes possible to enhance community effectiveness and prevent the emergence of emotional difficulties which may require traditional psychiatric intervention.

The principle modalities used at Keesler are consultation, education, and experiential learning. Education requires communication of human relations concepts and the early identification of student problems. Experiential learning requires group encounters designed to increase social system skills and enhance humanistically oriented leadership skills. Emphasis is placed on the use of newly acquired skills in the work setting. Effective consultation is a crucial component because it ensures "client" involvement and support of any HDC activity. HDC cooperation with the training community is enhanced by the assignment of HDC consultants to each training squadron. All of the programs mentioned below were planned with the participation of training personnel.

In January, 1979, the Keesler Mental Health Department was assigned additional professional staff, and officially christened its new primary prevention program the Human Development Center (HDC). Keesler's approach for conducting a program to work at supraindividual levels evolved into four general phases: (1) Orientation of the training community to the pro-

PRECEDING PAGE BLANK-NOT FILMED

gram (2) Identification of problems and trends in the training community (3) Interventions geared for supraindividual levels and primary prevention, and (4) Process review. Brief descriptions follow of some activities designed to implement and integrate these phases into a consistent program for support of the training mission.

Orientation. Group Commanders and Deputy Commanders for Training are routinely informed of HDC's operation. Students newly arrived from Lackland are briefed every day in the Mission Application Seminar (MAS) about HDC resources. In addition, new student officers are informed about HDC in their own MAS briefings.

Problem Identification. There are five major ways that HDC is attempting to identify problem trends in the training community. (a) Participation on one of the Base Commander's consulting groups, which identifies problem trends and suggests remedial actions. (b) Participation on a "student awareness council" which airs problems of new students. (c) Nominal groups-a technique for prioratizing personal and organizational problems. (d) Collection of training data such as the number of school eliminations and the number of disciplinary actions. (e) The use of a "Student Problem Rating Form" which doubles as a facilitator of intake interviews and a research instrument for the differintiation of problems by squadrons and AFSCs.

Interventions. There are six major intervention categories. (a) Consultations. HDC staff regularly consult with schools and squadrons on adjustment to the training experience. Air Traffic Control and Morse System students are visited every week to discuss coping mechanisms. Staff consult with squadrons about HDC programs and the counseling and disposition of students. (b) Instruction. Instruction is given at an Academic Counseling Course attended by most technical instructors. The course is offered by TIC, and imparts client-centered counseling skills. In addition, HDC staff teach Transactional Analysis and interpersonal communication techniques at the First Sergeants Academy. (c) Communication Skills. Junior Officers from the Com-Electronics Officer Training School attend workshops on communication skills for the facilitation of management techniques. A Peer Counseling Program trains students in client-centered counseling techniques. These students are introduced to their squadrons, counsel peers with minor adjustment problems, and are liasons to HDC for early identification of more serious problems. A Creative Listening Workshop was conducted with squadron NCOs to impart client-centered techniques. (d) Management Skills. These programs rely primarily on Human Relations training. A Leadership Effectiveness Training Laboratory is conducted for students. The lab emphasizes experiential learning in trust, group process, and problem solving in groups. A similar experience is being planned for personnel in the Com-Electronics Officer Training School. An Organizational Development workshop for technical instructors has been successfully completed. (e) Personal Growth - Adjustment Groups. A student group emphasizing personal growth is conducted every week. Adjustment Groups for students with training-related difficulties are offered on an as-needed basis. (f) Individual Counseling. Students are counseled on an individual basis for training-related adjustment difficulties, often referred by training squadrons. Students who need more concentrated psychotherapy are

referred to Keesler's Outpatient Mental Health Clinic.

Process Review. The process review phase of HDC activities have gained added impetus since October, 1979, when the ATC Surgeon General directed studies to assess HDC effectiveness. A research paradigm (Lagarde, Jenkins, Patrissi, and Stokes, 1980) outlines the analysis of community base-line data and a structure for the assessment of specific programs at ATC bases.

Following are some studies in progress. (a) A combined study of the three active HDC programs will determine if there have been appreciable effects on training related parameters such as discharges and school eliminations. (b) The first analysis of the Student Problem Rating Form confirms that it does discriminate among squadrons and AFSCs on some problems. It does not seem to discriminate well on self-perceptions of students. Sheppard AFB is also collecting data on this form. (c) A study of the Leadership Effectiveness Training Lab showed that students changed on their perceptions of how often they applied positive human relations behaviors, particularly in Communication and Social Relationship Skills. (d) Sheppard AFB is conducting studies to assess the effectiveness of HDC teaching techniques and the effectiveness of consultants to training squadrons. (e) Data on training community parameters will be charted as indices of problems in the training environment.

Summary

The HDC program was started to enhance the ATC training mission through increased sensitivity to supraindividual processes in the training community. Professional consultation, education, and experiential learning activities were discussed as implementations of HDC's primary prevention mission. For example, activities such as Peer Counseling focus on interpersonal communication skills. Leadership Effectiveness Training focuses on the supraindividual levels of the group and organization. Process review and consultations are designed to continue vital feedback on HDC effectiveness and mutually supportive liason between HDC and the training community.

References

Angyl, A. Foundations for a Science of Personality. New York: Commonwealth Fund, 1941.

Glad, D. Operational Values in Psychotheropy. New York: Oxford University Press, 1959.

Lagarde, H., Jenkins, J., Patrissi, G., and Stokes, J. A Research Paradigm for an Air Force Community - Primary Prevention Program, and a Proposal for the Analysis of Community Base-Line Data. Unpublished proposal submitted to the ATC Surgeon General, Randolph AFB, Texas, 1980.

Milsum, J. The hierarchical basis of general living systems, in Trends in General Systems Theory by Klir, G. J. (ed.). New York: Wiley, 1972, 145-187.

von Bertalanffy, L. General Systems Theory. Human Relations, Braziller, New York, 1968.

"The Human Resources Test and Evaluation System (HRTES): Its Concept and Design" [1,2]

J.D. Kaplan and W.H. Crooks
Perceptronics, Inc.

ABSTRACT

PROBLEM -- The ability of a military system to carry out successfully the missions for which it was designed is dependent upon the performance of its operators and maintainers. Current procedures of test and evaluation systems give relatively little consideration to the importance of human performance in system evaluation. How can test planners and evaluators decide what human performance to measure, how to measure it, how to evaluate measure outcomes, and how to determine the causes of inadequate outcomes?

RESULTS -- Through an analysis of existing approaches to testing, a survey of weapon system experts, observation of operational tests, and the utilization of decision making theory, a concept was developed for a system to aid the testing community.

The result of this concept is the creation of a two-volume handbook which is designed to comprehensively aid its users in the writing of test plans, the evaluation of test results, and the diagnosis of probable causes of human performance inadequacies. This two-volume handbook is called the Human Resources Test and Evaluation System (HRTES).

1.0 Introduction.

Testing and evaluation can be defined as the art and science of asking those questions, the answers to which make a difference. Implicit in this statement is the recognition that evaluation is intended to enlighten decision makers by obtaining the data with which they can identify how actual systems deviate from desired or ideal states. Answers to evaluation questions rely, therefore, on well-formulated statements of expectations (anticipated performance and performance outcomes), the specification of relevant performance criteria and associated standards, and the adequacy of procedures used to obtain necessary data by which expectations can be compared with performance. As the number of interactions within operational tests increase, the complexity of evaluation tasks becomes greater; but as systems become more expensive, the necessity for meaningful testing and evaluation also grows larger.

---

[1]This research was supported by the U.S. Army Research Institute for the Behavioral and Social Sciences under Contract No. DAHC19-77-C-0055.

[2]The views expressed in this paper are those of the authors and do not necessarily reflect the views of the U.S. Army or of the U.S. Department of Defense.

PRECEDING PAGE BLANK-NOT FILMED

Increasing weapons system complexity, the higher costs of materiel and personnel, and increasing demands for highly trained operators and maintainers have placed heavy demands on materiel testing programs. The increased life cycle costs from untimely detection of design errors, coupled with the knowledge that approximately 50 to 70 percent of all failures of major weapons and space systems are caused by what appear to be human-initiated failures, underscores the importance of including human resource considerations throughout the system acquisition process.

Thus, recent Army Regulations (AR 1000-2, Operating Policies with Systems Acquisition by the Department of the Army) state that "Organizations having logistic and user responsibilities must become involved in the program at an early time, in order to be prepared to carry out their responsibilities in parallel with those of the developer during the fullscale engineering development phase."

Add to the foregoing the realization that testing and evaluation has characteristically given major emphasis to the evaluation of the hardware components of systems with relatively minor consideration of their human components. Consideration of the human oriented aspects of man-machine systems (training, Human Factors Engineering leading to increased operability and maintainability, and operator/maintainer selection) are often relegated to secondary roles. Frequently, systems are tested and evaluated without human performance criteria, and according to individual and unspecified methods. Each test specialist involved attempts to apply the best professional standards and procedures. However, no complete systematic guidelines for human component oriented testing and evaluation are available to focus the test planning and evaluation efforts in a detailed fashion.

1.1 Problems

There are three general problems found in the assessment of the role of human performance in man-machine systems. The first of these is the problem of relevance. It is impressive to take measures of human performance and system design, but such measures have utility only insofar as they relate to the ultimate purpose(s) of the system being tested. That is, systems, theoretically, have been designed and constructed for some tactical or strategic purpose. Human operators and maintainers are a necessary part of the system, without which such purposes cannot be fulfilled. Therefore, measures of human performance and measures of those aspects of system design which interact with humans ought to be directly related to those ultimate purposes which mandated the creation of the system.

The second general problem of assessment is that of resource limitation. Characteristically, testing is underfunded, considering the potential importance of its outputs, and the most underfunded component of testing is, usually, human performance testing. The effect of the constraints placed on such testing by lack of resources is the necessity for measuring only those items which are absolutely vital.

The third general problem is the assignment of meaning to data collected. Data have meaning when they are compared to some criteria. Their utility is considerably impaired without such comparisons. The criteria which may be used as a template for data fall into categories based on their origin. These points of origin are:

(1) Equivalent, competing system data
(2) Historical data from predecessor system
(3) Research and development data
(4) Data from authority (either technical experts, tactical/strategic planners, or managerial experts).

There are two major objectives in assigning meaning to such data. The first is to determine the required level of performance of a system and, therefore, its potential acceptability. The second is to determine probable areas of significant system deficiency so that possible solutions and tradeoffs leading to such solutions can be logically considered.

2.0 Results

The Human Resources Test and Evaluation System (HRTES) is an attempt to provide some aid in these three problem areas. The following sections will present capsule descriptions of the concepts developed to deal with these problems.

2.1 The Problem of Relevance.

The HRTES concept views the understanding of the ultimate purposes of a system as being basic to the measurement of the system in an operational test. These ultimate purposes are referred to as System Functions to avoid introducing the existing fuzziness in the concept of "missions." The difficulty inherent in this sort of evaluative structure is the manner in which the connection is made between a System Functions and the measures of human performance. It is this connection which establishes the relevance of the measures. That is, the more coherent and logical the connection between System Functions and measures, the more relevant are those measures. In HRTES, this linking is accomplished by first defining systems according to their generic class. Each generic class is defined by its general functional and hardware similarities.

The second stage in the linking process is the assignment of System Functions to each generic class.

In the third stage of the linking process, each of these System Functions is decomposed into those elements which are required, of the system, to fulfill that System Function. HRTES refers to such elements as System Performance Issues.

These SPI's may, in turn, be linked with environmental, tactical, target, personnel, and/or operational conditions to produce more specific issues. The answers to such System Performance Issues determine the ability of the overall system to perform its System Function. The system's failure to

reach criterion in one of these areas may be caused by hardware, training, Human Factors Engineering, human selection, or a combination of several of these factors within the areas of system operation and/or maintenance.

In the fourth stage of the linking process, each SPI is decomposed into Human Performance Functions (HPF's) which are required to carry out the SPI. HPF's are a form of task which has been so defined as to be generalizable to different, but functionally related systems. HPF's are collected in HPF-Groups. An HPF-Group consists of all those HPF's needed to perform either the operational or maintenance portion of an SPI.

In the fifth stage of the linking process performance measures and criteria are attached to each HPF. HRTES views performance measures as falling into three categories: measures of time; measures of accuracy (made through measuring and counting errors); and a combination of time and accuracy. If an HPF is not performed according to its criterion, this inadequacy may have been produced by inadequate hardware performance, Human Factors Engineering, training, and/or selection of the human who performed the HPF. Required data collection for diagnostic measures of these areas are also specified at this stage of the linking process. Thus, systems have System Functions; System Functions have System Performance Issues; System Performance Issues have Human Performance Functions (collected in Operational and Maintenance HPF-Groups); Human Performance Functions have performance and diagnostic measures and criteria. Each system is linked to its relevant performance and diagnostic measures of human performance, and the linkage between systems and measures is complete. Unfortunately, the linking concept tends to produce an exponential growth as it fans out from generic classes to measures, necessitating some form of limitation.

2.2 The Problem of Resource Limitations.

The problem of the limitation of testing resources is addressed through a determination of the relative criticality of the elements in each layer of the HRTES hierarchical model. Each layer of the hierarchy, linking generic classes to measures, consists of competing elements. Each element, in a given hierarchical layer, has a relative level of criticality in reference to one element in the preceding layer. That is, some System Performance Issues are more critical than are others in fulfilling the parent System Function. This technique permits continuous elimination of noncritical elements to be tested, thus reducing the total number of human performance measures to be considered. Thus, the result is a limited number of critical human performance measures, directly related to the system, and an audit trial which specifies why each measure was related.

2.3 The Problem of Meaning.

Data, by themselves, have no meaning. Meaning must be assigned to them on some basis. Ideally, data could be assigned meaning by comparing them to validated, universal criteria. Unfortunately, there are not many universal criteria available, and those that one finds, in the literature, are frequently of questionable validity. HRTES provides a method for:

assigning meaning to the data obtained by human performance measures, determining which Human Performance Functions are significant and must be diagnosed, and diagnosing the probable causes of the inadequate performance of the significant Human Performance Functions.

HRTES assigns meaning to performance data through the use of Value Functions and a hierarchical model of system activities called an Evaluation Tree. A Value Function is a graphical representation of the values given to various possible outcomes of a performance measure of a given Human Performance Function by one or more experts. Through the use of Value Functions, HRTES converts all actual performance measure outcomes to experts' values of those outcomes on a common scale of value. The same procedure is also followed for the performance criteria of all Human Performance Functions. These values of measured performance and performance criteria are then applied to the Evaluation Tree through the procedure of "folding back" the Tree.

The Evaluation Tree is a hierarchical structure of six levels of nodes, at each level connected by branches to certain nodes in the adjacent levels. The six levels are: System Level, System Function Level, System Performance Issue Level, Human Performance Function-Group Level, Human Performance Function Level, and Statistic Level. The nodes at each level have a criticality weight based on the information required to select them. The process of "folding back" the Tree results in performance and criterion values for each node of the Tree, taking into account the values generated by the Value Functions and the criticality weights. Thus, human performance data are assigned meaning since one can now determine both their value and adequacy, in terms of the criterion, for each node of the Evaluation Tree.

The completed Evaluation Tree can then be used to determine which Human Performance Functions should be diagnosed. There is little purpose in diagnosing the cause(s) of acceptable performance, nor the cause(s) of relatively noncritical performance. By examining the nodes of the Evaluation Tree, one can determine which are both critical and unacceptable in terms of their performance. Such nodes can then be traced, through the branching structure, down to the Human Performance Function Level. Those Human Performance Functions which are thus identified are diagnosed, and all others left alone, thereby reducing the size of the diagnosis procedure.

Inadequate performance may be caused by some inadequacy in the performance of the system's hardware, poor training, inferior Human Factors Engineering, and/or inappropriate human selection. HRTES aids the process of diagnosing probable causes of inadequate performance by providing its users with both a series of measures and a method for systematizing the use of existing standards. The end products of the HRTES diagnosis are a number of hierarchically organized Indices of Adequacy of Training, Human Factors Engineering and Selection ranging from 100, for complete adequacy of each Human Resource Area, to zero. In this way, the three Human Resource Areas can be compared for adequacy in their contributions to the performance of each Human Performance Function. When a Human Resource Area is selected as being a probable cause of inadequate performance, it can then be decomposed to

more specific Indices, etc. So, probable cause(s) of inadequate performance can be stated at whatever level of specificity required for the particular purpose and audience.

3.0 Conclusion.

Using the Human Resources Test and Evaluation System (HRTES), it should be possible for test planners and evaluators to determine exactly: which items of human performance (of system operators and maintainers) to measure; which measures to use for these items of human performance; what criteria to use for the measures; how to design the test; and what resource measures to take during the test. Once a field test is completed, it should be possible to use HRTES to determine: the level of human performance in the system being tested; the effects of human performance on the system; the critical items of human performance which produced system performance inadequacies; and the probable causes of inadequate human performance.

Evaluation of an Image Quality Assessment Technique Based on Magnification

Gilbert G. Kuperman
Air Force Aerospace Medical Research Laboratory
Wright-Patterson Air Force Base, Ohio

Abstract

Resolving Power (RP) measurements have been the mainstay of photographic image quality assessment for over forty years. Nominally "better" techniques require the application of microdensitometers and digital computers. This research investigated the merits of the Visual Image Evaluation (VIE) technique when applied to high resolution camera system imagery. The photography included the degrading effects of target contrast and one-dimensional sinusoidal vibration. VIE was found to correlate well ($r = 0.9492$) against RP when both techniques were applied by trained image analysts.

Photographic acquisition systems continue to be one of the major sources of information needed to satisfy intelligence and mapping requirements. The Air Force has instituted a quality assurance program, based on image quality assessment methods, to specify the operational performance of these collection assets. The requirements for image assessment range from determining if an acquisition system operated satisfactorily with respect to design specifications, to determining the causes of evident degradation so that corrective action may be initiated. The first case might require the use of primarily objective, analytical techniques which exploit highly specialized mensuration devices, viewers, microdensitometers, and computers, while the second problem could be satisfied on the basis of subjective methods, requiring little specialized equipment.

The most commonly applied and best understood image assessment method is the estimation of system resolving power (RP) obtained by reading photographs of tribar targets. The tribar target consists of patterns of triplets of rectangular bars of specified bar width, bar length, interbar and interpattern separation, and bar-to-surround reflectance contrast. The bar width changes between successive patterns in a specified progression (typically by the sixth-root-of-two), providing a range of sizes. The unit to which tribar targets are designed and in which RP values are reported is the number of cycles per millimeter (cy/mm) represented by the smallest pattern judged to be resolved in the photographic image. A cycle is the combined bar width and space for that imaged pattern.

The Defense Intelligence Agency (DIA) has developed (in 1975) a standard procedure for performing RP assessments. This procedure includes the training and certification of RP readers through the demonstration of criterion performance.

The Visual Image Evaluation (VIE) technique is intended to provide an alternative to the use of RP and does not require imaging of tribar targets. It is based on the observation that photographic grain and other physical characteristics place a limitation on the maximum optical magnifi-

cation achievable in photographic enlargement and printing. Gliatti (1978) describes obtaining VIE readings as follows:

> "The procedure used is to view the imagery using a variable magnification zoom binocular microscope. The imagery is magnified to a point of empty magnification; i.e., where no additional information is obtained with increased magnification; in fact, information is lost due to grain, limited image resolution, motions during exposure, image contrast and/or other degrading factors."

The simplicity of the VIE technique is a compelling argument for its operational application. A search of the literature, however, yielded little encouragement as to its robustness. HRB-Singer Inc., for example, performed a limited study of the utility and robustness of the VIE method for the Rome Air Development Center in 1971. Only five frames of aerial photography, exhibiting RP values in the range of 7.3 through 31.3 cy/mm, were used in the study. No significant correlation was found between the VIE readings and the RP references. The Tactical Air Warfare Center also performed an evaluation of VIE and reported negative results in the 1972 Constant Quality final report as follows:

> "No correlation ... could be established that would be useful to tactical image quality analysis. Individual differences compounded by contrast variance, scale variance, and target type are critical problems which could not be adequately defined or categorized using typical tactical imagery."

The subjects used in the above referenced studies were trained and experienced tactical photographic interpreters, accustomed to performing detection, classification, recognition, identification, plotting, and mensuration tasks. Experience with VIE at the Air Force Sensor Evaluation Center, which uses both RP and VIE as quality assurance measures, suggested that VIE could be successfully applied by image quality analysts. The present research is an investigation of the utility of the VIE technique against the established RP procedure already in common use.

Method

Subjects. Four male subjects were used in this study. Each demonstrated at least 20/20 Snellen visual acuity (corrected or uncorrected). All were employed in a government laboratory in positions requiring them to perform subjective estimates of photographic image quality as a primary job function. All subjects had been certified, by the DIA procedure, as RP readers. All subjects had received on-the-job training in the application of VIE and had accumulated at least six months experience in using it as a part of their routine, daily duties.

Photography. A special purpose photographic system, described by Sun (1967), was developed for the creation of motion-degraded imagery. Tribar targets, at modulations of 0.76, 0.66, 0.52, and 0.08, were imaged onto Eastman Kodak Type 3404 High Definition Aerial film using a 250 millisecond exposure time. A 30 Hz sinusoidal motion was applied to the film plane during each exposure. The peak-to-peak amplitude of the motion was set at one of seven levels: 0.00000 (static), 0.00025, 0.00050, 0.00075, 0.00100, 0.00125, and 0.00150 inches. Four replicate photographs were

obtained at each of the twenty-eight Target × Amplitude conditions.

Apparatus. The light table and viewing equipment used for both RP and VIE readings were in accordance with the specifications presented in the DIA procedure. The light table used, which includes the light source, was a Richards Corporation Master Interpretation Module System self-standing, elevating table, equipped with a vacuum stage. The table was equipped with a Bausch and Lomb StereoZoom 7 Power Pod which was used with 10X wide field eyepieces and a 2X supplementary lens attachment.

Procedure. All subjects read all imagery under both techniques. The research imagery was intermixed with other photography being assessed by the subjects as part of laboratory or flight test evaluations of photographic system performance. All imagery used in this experiment was in the form of original negative. The subject placed the imagery on the light table and adjusted the illumination level, focus, and magnification as required. He then recorded the appropriate reading manually.

Results

The data produced by the RP method were employed as a baseline against which to compare the VIE estimates. A within-subjects, repeated measures, analysis of variance (ANOVA), BMD08V (Dixon, 1977), was applied to the RP data. Targets, Amplitudes, and the Target × Amplitude interaction were all found to be highly significant ($p<0.01$). The same ANOVA was applied to the VIE data. Subjects, Targets, Amplitudes, Subjects × Targets, and Subjects × Amplitudes were found to be highly significant ($p<0.01$) and the Targets × Amplitudes interaction was found to be significant ($p<0.05$).

The highly significant Subject difference found in ANOVA for VIE merited particular attention. Tukey's HSD statistic (Roscoe, 1975) was applied to the VIE data. One pair of subjects were found to be significantly different ($p<0.05$) from each other and all other pairings produced highly significant ($p<0.01$) differences.

An Omega-squared test was also applied to each data set (Simons, 1971). The independent variables and their interactions accounted for 96.5 percent of the total experimental variance with respect to RP and 87.4 percent with respect to VIE.

A linear regression analysis (BMD03R, Dixon, 1977) was carried out to compare RP and VIE directly. They were found to be highly correlated variables, exhibiting a correlation coefficient of $r = 0.9492$. The regression equation for the line of best fit was found to be:

$$VIE = 0.527\ (RP) + 13.4.$$

By inverting this equation, VIE readings, in diameters of optical magnification, can be used to produce estimates of resolving power, in cy/mm, at least over the ranges of conditions employed in this study. The inverted equation has the form:

$$RP' = 1.748\ (VIE) - 23.4.$$

Discussion

Both techniques appear to be extremely responsive to changes in photographic image quality. This is supported by the unusually high percentages of total experimental variance accounted for in the ANOVAs (as estimated by the Omega-squared test).

The VIE technique is an excellent correlate of RP and the relationship between the two methods is highly linear. This is borne out by the regression analysis.

The VIE technique is susceptible to differences between readers. The ANOVA for VIE showed the main effect of Subjects and all second order interactions which included Subjects to be highly significant. Tukey's HSD test, further, showed all subjects to be significantly different from each other.

Conclusions

Under the conditions of target contrast and image motion employed in this study, the VIE technique was found to be almost equal to RP as a means of measuring image quality. VIE readings were found to be easily converted into cy/mm, the unit used in reporting image quality for quality assurance purposes. The individual differences manifested between subjects is disturbing; perhaps a formal training and criterion performance procedure, as exists for RP readings, can be developed which would at least ameliorate this circumstance. Based on the results of this research, the operational utilization of the VIE technique for photographic system image quality assessment is recommended.

References

Defense Intelligence Agency. Standardized assessment and expression of tribar resolution. Washington, D.C.: Author, 1975.

Gliatti, Edward L. Image Evaluation Methods. Proceedings of the Society of Photo-Optical Instrumentation Engineers, 1978, 137.

HRB-Singer Inc. Preliminary examination of a magnification concept for evaluating photographic imagery. Rome, N.Y.: Author, 19 January 1971.

Tactical Air Warfare Center. Constant quality. Eglin Air Force Base, Fla.: Author, September 1972.

Sun, Y.C. Advanced photo subsystem component analysis: Effects of vibration and other random motions on high resolution photography (sinusoidal vibration study and experimentation phase) (AFAL-TR-67-260). Wright-Patterson Air Force Base, Ohio: Air Force Avionics Laboratory, December 1967. (NTIS No. AD-843 786)

Dixon, W.J. (Editor) BMD: biomedical computer programs. Berkeley, Ca.: University of California Press, 1977.

Simons, Charles N. Considerations for the proper design and interpretation of human engineering experiments (P73-325). Culver City, Ca.: Hughes Aircraft Company, December 1971.

Roscoe, John T. Fundamental research statistics for the behavioral sciences (2nd ed.). New York: Holt, Rinehart, and Winston, 1975.

A Human Operator Model for Command and Control Evaluations

Charles C. Jorgensen
U.S. Army Research Institute for the Behavioral and Social Sciences
Fort Bliss, Texas 79916

Abstract

The development of automated air defense weapons has sensitized the Army to the need for assessment of human performance contributions in linked command and control systems. One such system which operates at both the group and battalion levels is the AN/TSQ-73 Missile Minder. Because the state of the art in human simulation techniques has developed rapidly in the last five years, particularly in network and task description languages, a simulation model of the AN/TSQ-73 was developed to assess the feasibility of these methods for human performance evaluation. The emphasis was on a psychological model of a weapon assignment operator including task interaction with the battalion and Missile Minder tactical software system. Simulation required utilization of task descriptions, sub-models of physical hardware, external fire units, radar information, and operator psychological capabilities. The simulation is currently being validated through comparison with live operator performance within tactically configured AN/TSQ-73 hardware.

The Army is in the process of fielding an advanced series of mobile automatic data processing command and control systems such as the AN/TSQ-73 missile minder. The AN/TSQ-73 is designed as a group or batallion system for Nike-Hercules and Hawk air defense units. Air defense coordination is accomplished through integration of radar and identification friend or foe (IFF) data from local and remote sources via console display. Programing of the automatic data processing equipment generates alphanumerics, track and site symbols, map symbols, coordinates, and air corridors. These symbols are assigned to radar and IFF information to provide an operator with a display of aircraft and missile targets. Target data, fire unit profile data, and defended point characteristics are processed and analyzed automatically for primary and secondary fire unit selection and weapon assignment. Tactical operation of the system is accomplished by one officer and two operator/repairmen. They may perform their duties in a variety of modes depending upon the type of air threat, doctrine, or hardware conditions. These modes include fully automatic, manual, or mixed man/machine tasks. Examples of the modes include automatic or sector scans for air track identification, track hooking procedures such as sequential, position, or numeric methods, and automatic or manual fire unit selection and weapon assignment.

Because the Missile Minder is only one part of a complex net of low, medium and high altitude weapon and communication systems, the instantaneous load on a given part of the air defense net becomes extremely difficult to assess. Full scale field exercises are too expensive and detailed sensitivity analysis of the enormous number of possible situations

make decision-making a challenging problem. Although the inclusion of live operators within system simulators can provide test data during training, the inputs to those trainers are often artificially restricted by the lack of information about dynamic changes in the battlefield environment as well as human operator behavior in other weapons. Although many simulation techniques exist for characterization of hardware and highly predictable parametric phenomena, only recently have the tools been developed which permit rapid modeling of man/machine tasks. Examples include the Navy Human Operator Simulator (HOS) and the air force System Analysis of Integrated Networks of Tasks (SAINT). Little work has been done which uses these languages to encode psychological insights about an operator into usable modules for complex evaluations in command and control.

The AN/TSQ-73 Model Generation

Recognizing the need for human operator models, the Army Research Institute began an in-house effort aimed at selecting operator modeling methodologies in 1977. The result of this preliminary effort was the selection of a specific problem area (the AN/TSQ-73 operator/repairman), a command and control system, (the Missile Minder) and an appropriate analytical language for simulation (SAINT). SAINT utilizes a graphic approach to modeling in which the system to be analyzed is represented by a network using a special symbol set. The fundamental elements of a SAINT network are tasks, resources required to perform the tasks (equipment and/or personnel), relationships among tasks, and system status variables (state variables). System performance is related to which tasks are performed, the manner in which they are realized, and the extent that states of the system are achieved and maintained. In addition to providing a network model, the approach allows specification of conditions and constraints. Once a network model of a system has been developed, the SAINT program automatically generates system performance estimates through analysis of network properties. It is this capability which made the language particularly valuable for evaluation of human operator behavior within the AN/TSQ-73.

The First Year Model

During 1978, a preliminary model was developed to test the feasibility of the approach (Wortman, Hixon, and Jorgensen 1979). This model included representations of each of the primary elements surrounding an AN/TSQ-73 operator as well as detailing basic task functions to be performed Included in this model were submodels of the air threat, the Missile Minder threat and evaluation software, fire unit behavior, missile characteristics, timing constraints, and the task elements comprising the job of weapon assignment (Hixon and Wortman 1979). Although providing significant analytical capabilities through scenario evaluation options, the initial model minimized the attention given to human psychological processes in favor of developing a realistic system environment. Human performance characteristics such as radar scope symbol selection were included only in general probabilities based on the type of threat information. For example, standard times were associated with selection of hostile track symbols and these times were generally faster than those associated with

friendly or neutral tracks. Similar simplifying assumptions were made for idle time behavior, or communication.

The Second Year Model

In 1979, a second model was begun which has the operator processes as its main focus. The expanded model may be logically divided into five functional areas:

1. Operator tasks
2. Fire unit processing procedures
3. Aircraft control procedures
4. Operator visual search activities
5. Psychological process models

Operator tasks are modeled with a high degree of detail and reflect individual steps taken by an operator to perform duties. Activities always initiate with the operator searching the AN/TSQ-73 radar scope for a specific symbol to process. Once the operator has found an item on the screen for attention, the state description representing the operator is moved to the portion of the task network corresponding to the performance of that activity. After completing the activity, the operator is returned to the original search mode to process another track. There are exceptions to this cycle behavior which can interrupt performance. Ignoring these cases for the moment, the operator activities fall into nine possible categories. These are as follows. The first is the SEARCH activity which includes the operator scanning and selecting appropriate information for further processing. The second is the VIDEO TRACK activity in which decision making functions take place about how a particular track will be processed. Decisions are integrally tied to operating modes such as auto track initiation, sequence or position hooking, command corridors, or overides. The third is the OBSERVE UNKNOWN activity in which the operator determines the interrogation mode to be used, initiates identify friend or foe procedures, processes the result and takes appropriate actions. The fourth is the OBSERVE FRIEND activity in which the operator either returns to process other tracks or issues cease fire commands to fire units. The fifth is OBSERVE FIRE UNIT and is used to detect fire units that are out of action, drop the unit from the system software,and clear screen clutter. The sixth is OBSERVE HOSTILE in which the operator determines if the track will be immediately processed, assigns a fire unit, or modifies computer recommended battery choices. The seventh is OBSERVE SPECIAL which relates to clearing operations associated with effective kills or status changes. The eighth is IDLE TIME and is a timing function tabulating free periods during which other activities such as operator rest can take place. The last is COMMUNICATION TIME and is included to simulate delay and confusion through use of verbal communications that interrupt the operators work or thought.

Fire unit procedures include activities necessary to maintain a realistic environment for the operator. The level of detail is only that necessary to provide operator interactions and therefore represents an aggregation of fire unit functions. The functions modeled include

message reception, interpretations, track acquisition, handoff, missile firing, and evaluations of the success or failure of firing. The fire unit may initiate automatic firing of a second missile, receive secondary messages from the AN/TSQ-73 operator, receive messages from the AN/TSQ-73 computer, execute cease fires, or temporarily hold fires on a given target.

Aircraft procedures permit the model user complete control over aircraft flight paths and characteristics. Control includes direction, speed, rate of turn, time at which the system can perform identifications, and point of appearance on the scope. Arbitrarily large numbers of aircraft may be used and simulated.

Operator scan procedures include three submodels. The first is a model of the operator physically moving his eyes over the radar screen. A Markov state model is used in which the screen is divided into eight sectors with associated transition probabilities. The second is a model of symbol assessment based on a mental picture of the screen created by the operator as symbols are scanned and selected. Each symbol is given variable priorities based on track specific information. The third model consists of the selection of particular symbols to process from the mental picture. Thus operator perception is captured in the mental image of the screen versus the actual physical image being presented. Also associated with the latter model is the possibility of communication requirements based on the track selected.

Psychological models were developed for more general operator characteristics. In particular, research was surveyed to select functions reflecting fatigue-both short and long term, training effects, and stress. These functions may be used to modify operator performance times throughout the model and are not associated with any specific task function. They can be called into play through the crossing of state thresholds as well as through the modification of individual activity variables.

Fatigue and training are functions of simulation event time within a run and may be broken down into two phases. During the first there is no effect, i.e., as time passes, no operator task performance change is directly attributable to either. During the second phase, repetition of operator tasks begins to train the operator and increases performance by decreasing task times. Counterbalancing this gain in speed is an increase in fatigue during which reactions slow and performance deteriorates. Each function can be varied depending on scenario conditions and assumed training level within the operator MOS catagory.

Stress is a complex function based on a calculation of operator work load rather than time. As load increases beyond a given level, a negative reaction becomes dominate and the operator's performance is degraded directly.

Model Outputs

The output from a simulation run falls into three categories. First is an echo check of the model compilation in SAINT. Second is an echo

check of the AN/TSQ-73 mission input including such details as the aircraft scenarios, operating and policy modes, and fire unit capabilities. The third consists of a detailed mission record and a series of statistical reports. In regard to this third area, the analyst currently has two options which can be used to study mission performance. Performance may be examined at the task level where each time and activity may be observed for the entire mission on an event by event basis. Thus, an analysist could tell at exactly what time an observation of a hostile was taking place, what the status of all hostile, friendly and neutral aircraft were at that time, and exactly what the fire units were doing. At a more global level, the second option summarizes the behavior of general task areas. The summary includes histograms of task performance times over the entire mission; mean, standard deviations, and observation counts for each user selected activity; and overall summaries across multiple runs and scenarios. Also included are records of fire unit behavior over time, missile expenditures, and an experimental measure of operator effectiveness. The latter is defined as the percentage of time during an exercise when a weapon assignment operator has more than four hostile aircraft in the radar range of the AN/TSQ-73 and upon which no operator actions are being performed.

The above are merely exemplary of the kinds of output options available for model analysis. The network structure of the SAINT language permits considerable freedom in choosing analysis points, the kinds of reaction times and error data available, and the specific operator scenario variables which may be considered.

In order to more accurately parameterize the model, actual operator switch times taken from live exercises are being compared with the model outputs. It is anticipated that the comparison process will then permit the use of the model to test operational policies that would have required excessively large and costly numbers of test exercises had live personnel been used.

REFERENCES

Wortman, D.B. Hixon, A.F., and Jorgensen, C., A SAINT Model of the AN/TSQ-73 Guided Missile Air Defense System, ARI Research Memorandum in press, 1979.

Wortman, D.B., and Alonzo, F. Hixson, III, A SAINT Model of the AN/TSQ-73 Guided Missile Air Defense System - User's Guide, Pritsker and Associates for U.S. Army Research Institute for the Behavioral and Social Sciences, ARI Field Unit - Ft. Bliss, Texas, in press, 1979.

The Role of the Operator in Designing for Maintainability

James W. Kochevar and J. F. Brock
Honeywell Systems and Research Center
Minneapolis, Minnesota 55413

Abstract

A seldom studied aspect of system maintenance is the role of the operator. Although the operator is likely to be the person detecting a fault, little has been documented concerning the role he or she may play in predicting faults before they occur. Anecdotal evidence indicates that operators can in fact do this and suggests that many systems provide state information which, if correctly interpreted, can tell operators about the future state of the system. The present paper argues that study of system-status information which allows fault prediction is a desirable addition to the more traditional ways of thinking about design for maintainability. Fault prediction can be enhanced by incorporating these features in system designs.

The problem of maintaining complex military systems is persistent (e.g., Bond, 1970; Carhart, 1953) and remarkably intractible to solution (Brock, 1978). Several patterns in how maintenance problems have been addressed are discernible. Designing systems which are more easily maintained by maintenance personnel (e.g., Crawford and Altman, 1972); designing systems which require less maintenance (e.g., the introduction of solid state electronics into complex systems); designing more sophisticated test equipment which functions somewhat automatically; improving the training of maintenance personnel (e.g., King and Duva, 1975); and providing proceduralized job aids for maintenance technicians (e.g., Joyce, 1975).

This paper specifically addresses the issue of designing for maintainability. Traditionally, design of equipment for maintainability has been taken to mean that equipment be designed to minimize the following: need for maintenance, maintenance time and effort when it is required, and hazards to equipment and maintenance personnel.

In designing equipment for maintainability, the following design areas are usually considered: accessibility, packaging, connectors, conductors, fasteners, labels, maintenance displays, maintenance controls, maintenance support equipment, test equipment, maintenance procedures, and job aids (Crawford and Altman, 1972).

Maintainability design is based upon traditional notions of preventive and corrective maintenance and design concepts often reflect expected maintenance practices at the organizational, intermediate, and depot level. Emphasis is placed on designing for the maintenance technician.

A seldom-studied aspect of system maintenance is the role of the operator. Although the operator is likely the person who first detects a fault (i.e., the system malfunctions), little has been documented concerning the role he or she may play in preventing system faults. Anecdotal evidence indicates that pilots, for example, routinely inform maintenance personnel that an aircraft subsystem "does not feel right" and maintenance actions to

PRECEDING PAGE BLANK-NOT FILMED

prevent failure should be performed. In these instances, the operator anticipates a future failure and initiates preventive maintenance actions that might not otherwise have been performed in time to prevent system failure. Occurrences such as this suggest that systems provide state information which, if correctly interpreted, can tell operators about the future state of the system. If this is so, one should be able to determine what specific system-status information is most useful in predicting system faults before they occur.

The term "anticipatory maintenance" is used here to describe those status system monitoring functions an operator should perform to detect and correct system faults before they occur. The term was selected to help distinguish this sort of maintenance activity from preventive or corrective maintenance.

Anticipatory maintenance differs from the usual concepts of preventive and corrective maintenance in several respects: frequency of occurrance, initiating conditions, performing personnel, and typical tasks. Table 1 summarizes some of the differences between preventive, corrective, and anticipatory maintenance.

Table 1. Characteristics of Three Maintenance Activities

	Preventive	Corrective	Anticipatory
Frequency	Periodic	As Needed	Continuous
Initiating Conditions	Planned	Reactive	During Operation
Performed by	Maintenance Technician	Maintenance Technician	Operator
Typical Tasks	Align	Fault Isolation	Monitor Displays
	Adjust	Remove	Compare Displays
	Inspect	Repair	Read Logs
	Lubricate	Replace	Predict Events
	Clean	Align	Communicate with Engineers
	Document	Adjust	Document
		Document	

The concept of anticipatory maintenance is sufficiently distinct from more traditional maintenance concepts so as to warrent separate consideration. Yet system designs rarely address anticipatory maintenance. The operator operates; the maintainer maintains. We are suggesting that those cues which lead the operator to make successful system predictions and intervene to prevent malfunctions are legitimate considerations for maintainability designs.

Although instances of design for anticipatory maintainability are rare, some do exist. For example, anticipatory maintenance has found acceptance in oil refineries and on board Navy ships in the monitoring of vibrations from pumps (Hanson, Note 1; DeLong, Note 2). Vibration sensors, coupled with appropriate processing and displays, allow process control operators to continuously monitor the status of critical pumps and anticipate their failure.

Anticipatory maintenance can often replace more traditional forms of maintenance. In the case of vibration monitoring in oil refineries, for example, regularly scheduled preventive maintenance on critical pumps is not possible. It is prohibitively costly to suspend processing while preventive maintenance is performed. Instead, vibration monitoring is used to help predict an upcoming failure in the systems. This ability to predict failure allows for scheduled shut down and corrective maintenance. The capability of scheduled shut down and corrective maintenance can have significant cost benefits. If the system were allowed to fail prior to provisions for corrective maintenance, the extent of required repairs might be greater and system down time might be longer.

In military systems, knowing that a component is approaching a failure window could allow for corrective measures being taken aperiodically but conveniently, e.g., while the ship is at pierside, the tank is in between missions, the aircraft is being refueled. Preventive maintenance is an attempt to cover all potential system malfunctions. But the best preventive maintenance system yet designed has failed to overcome Murphy's Law.

The idea of the operator playing a critical role in maintainability is generally accepted by engineering and maintenance personnel who have the idea explained to them (Corneilson, Note 3; Delong, Note 2). However, the systematic investigation of that role remains to be done.

Systems are designed with displays to provide state information, but rarely are these displays designed with the objective of detecting potential system failures. The human engineering task, therefore, is to supply this information to the operator without distracting from his or her primary task: operating the system.

The operator must perform a kind of trend analysis; displays which would provide the operator with historical data of a system parameter would aid such an analysis. On many systems, relationships among variables are accurate state indicators (e.g., heat, pressure, and flow rate; power input/output comparisons). Displaying appropriate values in proximity would enhance the operators grasp of relevant systems status.

The research problem is one of no data. Attaching the term "anticipatory maintenance" to what we have described has heuristic value. As stated above, people who should know are attracted to the idea. But there is no empirical evidence supporting the concept. Anecdotal evidence is hardly a

basis for changing design approaches to current or future systems.

A research program to identify techniques for exploiting anticipatory maintenance in system design is overdue. All three services can reap benefits from such a program--a program which would be remarkably inexpensive.

References

Bond, N.A., Jr. Some persistent myths about military electronics maintenance. Human Factors, 1970, 12 (3), 241-243.

Brock, J.F. Maintenance training and simulation: Design processes and evaluation. Proceedings 22nd Annual Human Factors Society Meeting, Santa Monica, Ca: Human Factors Society, 1978, 260-266.

Carhart, R.R. A survey of the current status of the electronic reliability problem. Santa Monica, Ca: The RAND Corporation, RM1131, August 1953.

Crawford, B.M. and Altman, J.W. Designing for maintainability. In H.P. VanCott and R.G. KinKade (Eds), Human Engineering Guide to Equipment Design (Revised Edition). Washington, D.C. : U.S. Government Printing Office, 1972.

Joyce, R.P. Performance aids and technical training: Complementary information sources, In W.J. King and J.S. Duva (Eds). New Concepts in Maintenance Trainers and Performance Aids (NTEC IH-255). Orlando, Fl: Naval Training Equipment Center, October 1975, 5-13.

King, W.J. and Duva, J.S. (Eds). New Concepts in maintenance trainers and performance aids (NTEC IH-255). Orlando, Fl: Naval Training Equipment Center, October 1975.

Reference Notes

Note 1. Hanson, T. Koch Refining Co. Visit. Honeywell Memo, 15 Aug. 1978.

Note 2. DeLong, J.L. Personal Communication with J.F. Brock, June 5, 1979.

Note 3. Corneilson, R. Personal Communications with J. F. Brock, June 8, 1979.

Improvements in US Army Rifle Marksmanship Based upon Recent Marksmanship Research

Seward Smith, Thomas J. Thompson, John C. Morey and Arthur D. Osborne
US Army Research Institute Field Unit, P.O. Box 2086
Fort Benning, Georgia 31905

Abstract

Because of concern about the quality of Army marksmanship, a major research effort in rifle marksmanship training has been conducted by the Army Research Institute Field Unit, Fort Benning, Georgia. Through field study, problems have been identified in instruction, course content and lack of shooting performance feedback. Results of three field experiments involving over 3,500 basic trainees at Fort Benning and Fort Jackson have resulted in implementation of a new basic rifle marksmanship training program. These experiments have tested and shown statistically reliable differences using improved zeroing techniques, detailed knowledge of results and simplified fundamentals. The paper will describe the problem analysis and experimental results including their implications for marksmanship training and equipment and for training effectiveness analyses in general.

Rifle marksmanship and its associated training and evaluation have had a long history of emphasis in the US Army. At least four major improvement efforts have been conducted within the last 25 years (e.g., Dees, Magner and McClusky, 1971; McFann, Hammes and Taylor, 1955).

The most recent of these research programs, and the subject of this paper, stems from rapidly growing costs of training and training equipment and from the Defense Department's greater emphasis on evaluating the products of training. In response to these pressures, the US Army Research Institute (ARI) Field Unit, Fort Benning, Georgia, began a systematic marksmanship research program in 1976. The purpose of this paper is to outline the directions and most promising findings of this research that have led to development of a new basic rifle marksmanship (BRM) training program.

Observations

The 1976 BRM Test

In 1976 a cost and training effectiveness analysis (CTEA) of marksmanship was conducted at Fort Jackson, S. C. The purpose was to see if cost/time savings could be realized by trying three alternate shorter programs in comparison with the Army Subject Schedule program (ASUBJSCD 23-72), the marksmanship program in use at the time. The four programs led to relatively similar final performance. The Army concluded that the shortest program of instruction (POI) was acceptably close to the baseline program and could therefore be substituted for it (TRASANA, 1977). This substitution was accomplished in April 1977 with the publication by the US Army Infantry School (USAIS) of a 37 hour, 334 round BRM POI (USAIS, 1977). It was to be an interim program pending further research development.

In retrospect two problems were evident in the 1976 test. The first was the assumption that the ASUBJSCD 23-72 program already led to

PRECEDING PAGE BLANK-NOT FILMED

acceptable BRM performance, but was just too costly. In fact, the assumption of adequate performance was probably unjustified. During the test that program produced only about 55 percent target hits in spite of nearly half of the target exposures occuring 50 to 150 meters from the firer. The second point was that the major reductions were in marksmanship fundamentals, zeroing and field firing - all areas which should be affording skill practice and therefore skill improvements. If repeated hours and rounds spent on these activities in the ASUBJSCD program were not leading to more firing skill, one would need to conclude either that skills were already maximized (which was unlikely) or that something was wrong with the training. It appeared wise, therefore, to conduct a careful analysis to determine the problems. This became an early, high priority ARI task.

BRM Problem Diagnosis

Four members of the ARI-Benning/Litton-Mellonics staff took part in BRM training at Fort Jackson, South Carolina, in March, 1978. By then the Infantry School's short BRM POI was in use. Early in that week of training it became clear what some of the major problems are.

Performance Feedback. The most serious of all problems is that the BRM program contains very little detailed feedback that can be used to sharpen marksmanship skill. At 25 meters fundamentals are compressed into 10 hours of formal training, during which time the trainee might fire as few as 27 shots. These 27 shots, although few in number, at least give performance feedback because the trainee walks down range to examine the 25 meter target after firing each three shots. From this point on, however, killable popup silhouette targets that fall when struck by a bullet are used for all remaining shooting (e.g., field firing, record firing). Unfortunately, the trainee has no way to determine where on the silhouette his hit is located and, in the event of a miss, may receive no cue at all where the bullet went.

It is a universally accepted principle of learning that there must be relevant, detailed and timely knowledge of performance results if successful learning is to take place. We found that the BRM program we participated in at Fort Jackson lacked most of that necessary feedback.

Instruction. The most serious instructional problem seems to be the general lack of knowledge and marksmanship teaching and diagnostic skill of the drill sergeants (who comprise the majority of the instructor pool). Very likely they have come through a similarly thin, non-instructional, non-feedback program themselves and have not seen first hand the down-range effects on bullets and have received little if any instruction or practice in how to teach BRM or in problem diagnosis and remediation. Due to budget constraints the Army has also greatly reduced the number of BRM teachers available so individual instruction as will as content is lacking.

Few performance measures are taken that could be used to check progress or that could be used to identify persons with problems. Also we noted that the standard zeroing target is confusing to use and zeroing in general is difficult to learn. Finally, it appeared that many trainees miss significant portions of the instruction and, because of the compressed schedule, have no opportunity to make up what was missed.

By the end of our visit at Fort Jackson as BRM training participants it was clear that there were several ways that the training could and

should be substantially improved.

We confirmed that our findings were accurate and general by going to several other training centers. We also visited Marine Corps training where we observed the effectiveness of good instruction and detailed performance feedback.

Experiments

Three major experiments were conducted to test promising ideas for improvement and to test candidate new instructional programs.

Experiment 1

The first experiment was conducted at Fort Benning with the US Army marksmanship unit (AMU) during September, 1978. It examined annual requalification of 82nd Airborne soldiers. Three experimental groups were tested. The first (n=89) went through one day of standard requalification with little instruction and little feedback. A second group (n=97) took part in a two day program with AMU instructors. The third group (n=88) was in a three day program that included extra AMU instruction and a day of shooting with accurate down-range feedback. Groups 1-3 averaged 23, 25 and 28 hits (of 40) respectively on record fire. Group three's 22 percent significant increase ($p<.001$) over the standard condition indicated the value of both qualified instructors and shooting performance feedback. For further information about this experiment see Evans et al, 1979.

Experiment 2

The second experiment was conducted at Fort Jackson, S. C., during October and November, 1978. A total of 2,124 basic trainees took part in an experiment designed to test use of a new rifle zeroing target and a period of added instruction during which the student fired at 75 and 175 meter targets and then walked down range to see the results. The new zeroing target is used at 25 meters. It gives simple step by step information to aid in making sight changes. It also gives information about the student's probable later down-range performance based upon the early use in training of this target. It therefore serves as a diagnostic tool to aid shooters who might later perform poorly at greater range if not corrected early in training.

Four groups were run: Standard training (22.7 hits); a group using the new zeroing target in otherwise standard training (23.7 hits); a group receiving down-range feedback training added to standard training (24.5 hits); and a final group given both the new target and down-range feedback (25.5 hits). The results showed a significant ($p<.005$) increase in record fire performance when the new zeroing target was used, a significant increase ($p<.001$) when down-range feedback was added and a yet greater increase ($p<.001$) when both were added to the training. It was concluded that the new zeroing target and down-range performance feedback would become parts of the projected new training. Details of this experiment are given in Smith et al, 1979.

Experiment 3

The final major experiment was also conducted at Fort Jackson. Data collection involved 1,151 basic trainees from April through May, 1979. During pilot investigations we had learned that, after zeroing the M16A1 rifle for 250 meters (the Army's battlesight zero), shifting to the long range rear sight (used to extend the rifle zero to be suitable for ranges in excess of 350 meters) causes the bullet strike at 25 meters to coincide

with actual point of aim. We therefore designed black on white silhouette targets for use at 25 meters to simulate (in visual angle) field fire targets at various ranges. Firing at these, the student could learn center of mass aiming with accurate feedback and instructors could help those students having problems before allowing them to move on to actual field fire popup target exercises.

These silhouette target exercises, along with use of the new zeroing target, down-range feedback and an attempt to bolster the instructor corps, became parts of candidate training programs tested in the experiment.

We again used a baseline control consisting of students given standard training. Three new programs were designed. All three used the new zeroing target and silhouette target exercises. Varied were number of instructors (standard vs. augmented staff) and use of the down-range feedback exercise vs. use of additional silhouette firing at 25 meters.

The results of the experiment are covered in detail elsewhere (Thompson et al, 1980). In general the group receiving all improvements (26.5 hits) outperformed the control condition (20.5 hits) by as much as 29 percent ($p<.001$). Those who missed some training were poorer (usually by 3 to 4 hits) than those who were present for all instruction ($p<.001$). In this experiment, as in previous ones, men scored higher than women ($p<.001$) by 1.8 to 2.6 hits depending upon group assignment. The use of more instructors yielded trends toward better performance but we quickly learned that having more instructors is not necessarily better if those added are not properly qualified to instruct.

The New Basic Rifle Marksmanship Training Program

Based upon these experiments and several other pilot studies, all performed in coordination with the US Army Infantry School, a new program of instruction has been developed and is currently in use at Fort Benning. It will shortly be exported to the field for general training use.

It is beyond the scope of this paper to detail the program. In summary, however, the program emphasizes five major points:

1. It stresses simplified fundamentals before moving on to field firing exercises.
2. It contains several diagnostic check points so that early problem detection and correction can occur.
3. It contains a natural progression from fundamentals through to rapid engagement of targets in combat-like setting with each exercise serving as a building block for those that follow.
4. It places major emphasis on feedback so students are given as much knowledge of their shooting performance as present technology and expense will permit. This is both for the poor shooter who needs mistakes corrected and for the good shooter to help him sharpen his skills.
5. Finally, the program is designed to aid the instructor to be a more adequate teacher. For example, an instructor's guide has been prepared and is currently being field tested.

Future Research

The ARI research team has now turned attention to the study of advanced individual and unit level marksmanship with a goal of developing a family of integrated training programs. There is a need to develop moving target training and to improve night and automatic fire shooting skills. Because target equipment plays a role in all of these activities

we have a major interest in and concern about marksmanship training devices.

References

Evans, K. L., Thompson, T. J., & Smith, S. FORSCOM/US Army Marksmanship Unit M16A1 rifle and .45 cal. pistol marksmanship training evaluation. US Army Research Institute, Draft Research Report, Feb., 1979.

Dees, J. W., Magner, G. J., & McCluskey, M. R. An experimental review of basic combat rifle marksmanship: MARKSMAN, Phase I. HumRRO, Technical Report 71-4, Mar., 1971.

McFann, H. H., Hammes, J. A., & Taylor, J. E. TRAINFIRE I: A new course in basic rifle marksmanship. HumRRO, Technical Report 22, Oct., 1955.

Smith, S., Thompson, T. J., Evans, K. L., Osborne, A. D., Maxey, J. L., & Morey, J. C. Effects of down-range feedback and the ARI zeroing target in rifle marksmanship training. US Army Research Institute, Draft Research Report, Feb., 1979.

Thompson, T. J., Smith, S., Morey, J. C., & Osborne, A. D. Effectiveness of improved basic rifle marksmanship training programs. US Army Research Institute, Draft Research Report, Jan., 1980.

TRASANA. Basic rifle marksmanship cost and training effectiveness analysis. Technical Report 16-77 (Volume II - Main Report), Sep., 1977.

US Army Infantry School (Directorate of Training Developments) M16A1 rifle marksmanship training program of instruction, April, 1977.

A COMPARATIVE SELF-CONCEPT ANALYSIS OF CROSS-CULTURAL MARRIAGE VARIABLES

Jaime I. Albornoz, Ph.D.
Clinical Psychologist
Mental Health Clinic
March AFB, CA 92518

The study examined whether or not differences existed in marriages between Caucasian Americans as compared to marriages between Latin Americans and Caucasian Americans. Variables analyzed were: self-concept congruency, marital adjustment, and reactions toward infidelity and double standard practices, at individual and group levels. Participating in the study were 128 subjects randomly selected throughout the San Diego area. Data analyzed revealed significant differences in self-concept congruency when comparing intracultural and intercultural groups. Conclusions were that cross-cultural self-concept congruence is not of the same degree, and that couples married interculturally may experience more difficulty in self-concept congruence attainment. Marital adjustment data revealed significant differences at the group level. Thus, cross-cultural marital adjustment differences between these groups is likely the result of cultural variables, rather than the direct result of personal differences, which may play a significant role in determining degrees of causes or failure in these types of marriages. Analysis concerning infidelity indicated a greater incidence on the males' part to engage in an extramarital affiar and to practice double standards. Reasons for engaging in this type of behavior, therefore, may be directly associated not to ethnocultural factors, but to changes in current social trends, mores and sexual freedom. Lastly, the study's findings revealed that males consistently ranked higher than females in areas of self-concept congruency, marital adjustment, and infidelity frequency.

STATEMENT OF THE PROBLEM: A need currently exists to examine cross-cultural aspects of marital relationships, i.e., the interactions of self-concept adjustment and infidelity issues as possible influential factors in what may constitute success or failure of marriages between Latin Americans and Caucasian Americans. Research into ethnic relations self-concept has revealed to a great extent attitudinal and perceptual reactions people have towards their own and other ethnic groups. According to a number of studies (Brand, Ruiz, and Padilla, 1974; Brigham, 1971), Brigham, in reference to individual attitudinal and perceptual reactions emphasized the point that underlying these reactions there is an assumed need to evaluate oneself and others and to perceive social categories in constructive and meaningful ways. The concerns of cultural problems and adjustment difficulties of cross-cultural marriages seems to have been little researched. According to Barron (1972), cross-cultural marriages have increased over the years. Research to date in the area of intermarriages, and most specifically, analysis of degrees of success or failure, does not reflect conclusive data of a significant nature, nor does it show a thorough sampling across the board. The subject of infidelity in marriage, its significance and possible effects in terms of marital success or failure, has been a topic of interest for social and behavioral scientists since the inception of marriage as an institution. Research on this topic in relation to cross-

PRECEDING PAGE BLANK-NOT FILMED

cultural marriages deals primarily with marriages among whites and blacks.
PURPOSE OF THE STUDY: The purpose of the present study was to determine whether significant differences were found within intracultural marriages as compared with intercultural marriages in terms of self-concept congruency, marital adjustment, and views towards infidelity. Specifically, the objectives of this study were to examine the following differences: 1. The differences in self-concepts among males and females in intracultural and intercultural marriages. 2. The differences in marital adjustment patterns among males and females in intracultural and intercultural marriages. 3. The differences of responses to questions pertaining to extramarital affairs. This study, particularly because of its uniqueness, sought to achieve findings which could assist other researchers in the field to develop new clinical approaches which would enable counselors to better understand the psychodynamics associated with these types of marriages.
PROCEDURE: The data for the study were compiled from 128 subjects randomly selected throughout the San Diego area. A cultural criterion was established in order to maintain the study's desired objectives within the prescribed scope. The intracultural group consisted of 64 Caucasian Americans (32 couples). The Latin American group consisted of 64 subjects (32 couples) formed by either Latin American males or females married to Caucasian Americans.
METHODOLOGY: Data for the study were gathered using three instruments: the demographic questionnaire (developed by the researcher) was used to gather information pertaining to subjects' interpersonal variables, and responses to questions regarding extramarital affairs. The second instrument was the Self-Concept Incongruency Scale (SCIS) developed by Weedman, Warren and Marx (1974). It was used to test hypotheses regarding self-concept congruency. The third instrument was the Short Marital Adjustment Test (SMAT), developed by Locke and Wallace (1959). It was utilized in this study to test hypotheses pertaining to marital adjustment.
SAMPLE CHARACTERISTICS: The interpersonal variables considered in the study revealed that the males in the intracultural group had a mean age of 34 years, and the females had a mean age of 29 years. In the intercultural group the mean age for the males was 32 years, in contrast to a mean age of 29 years for the females. The majority of couples in the intracultural group was in the $15,000 to $20,000 income bracket. In the intercultural group, the average income was slightly higher, approximating $20,000. The majority of couples in the intracultural marriages had been previously married, the length of their current marriage averaging 5 years. They had an average of two children, and possessed an education of 12-14 years. For the intercultural group almost 70% of the males and 50% of the females had been previously married. Most of the couples had been currently married for 3-5 years, had an average of approximately 1.5 children, and possessed an education averaging 15-16 years. The majority of the intracultural subjects was of the Protestant faith. In the intercultural group the majority of subjects was of the Catholic faith. In terms of occupations, almost half of the respondents in the intracultural group were in the trades, whereas slightly over half of the subjects in the intercultural group reported as being in the professions.
CONCLUSIONS AND IMPLICATIONS: When comparing the intracultural group with the intercultural group for self-concept congruence, significant differ-

ences were found at the .05 level of confidence. These differences can be attributed to cultural factors uniquely inherent in the respective cultures. This postulation is further supported by research into ethnic relations self-concept conducted by Brand, Ruiz and Padilla (1974) and Brigham (1971) which revealed that attitudinal and perceptural reactions people have towards themselves and other ethnic groups leads them to establish comparisons with their own and other groups in order to determine conceptual differences. It was important to note that when comparing the males of the two groups to the females, no significant differences were found. This could be largely due to the fact that approximately 80% of the men scored good self-concept congruence, as compared to approximately 64% of the females. This finding coincides with the results of field studies conducted by Weedman, Warren and Marx (1974), Weedman (1979), Marx (1976), Weisner (1976), which revealed that the lower the incongruency between the real vs. the ideal self, the better the self-concept congruencey and overall perception of the self. From aforementioned findings, the researcher concluded that cross-cultural self-concept congruence is not of the same degree when comparing groups and sex of respondents. It is furthermore concluded that in view of lower self-concept scores reported by females, especially those in the intercultural group, the implication is that perhaps couples married interculturally experience more difficulty in self-concept congruence attainment. An important consideration, however, is that the size of the sample was not sufficiently large to be able to make comparisons between the males and females of each group in order to predict directional outcome. This criterion is also true, and applies to the remaining variables under consideration in this study. The results of hypotheses pertaining to marital adjustment revealed that when comparing the two groups, significant differences existed at the .05 level of congruence, which is indicative that intergroup cultural variables may be the cause of these differences. This finding is supported by research studies (Bossard and Boll, 1957; Goldstein, 1942; Merton, 1941; Gordon, 1964) which conclusively revealed that differences in cross-cultural marriages are the byproduct of cultural orientations, and the system of social stratification in today's society. When comparing the males with the females of both groups, no significant differences were found in marital adjustment patterns across the board. This could be related to the fact that most couples reported good and average marital adjustment. Although significant statistical differences did not exist, an examination of marital adjustment results for intracultural and intercultural males and females, respectively, depicted some differences which may be, in the researchers opinion, related to culturally based personal adjustment patterns. This finding is supported by studies conducted by pioneer researchers such as Burgess and Cottrell (1939:10) who in referring to cross-cultural marital adjustment wrote, "If marriage has become a personal rather than a social relation, adjustment is to be defined in terms of personalities, their conflicts and accomodations, and the degree of assimilation taking place." It was concluded that the most important factors in marital adjustment were those of compatibility, affection, and social adaptibility. According to Walter Jr. (1952), the adjustment problem of the Latin American is mainly cultural. He postulated that differences in language, religion, and a marked difference in general attitude toward life sets them apart from other groups. Another

view supporting cross-cultural adjustment differences was proposed by Brewton (1951:75) who wrote: "The ethnic group is a human group bound together by ties of cultural homogeneity. Complete uniformity, of course, is not essential; but there does prevail in an ethnic group a high degree of loyalty and adherance to certain basic institutions, such as family patterns, religion and language." In view of the aforementioned findings, it can be concluded that marital adjustment differences among the intracultural and intercultural groups is a result of cultural variables, rather than as a direct result of personal differences among males and females of the two groups. The implication of these findings is that cross-cultural marital adjustment variables may play a significant role in determining degrees of success or failure in these types of marriages. The results of hypotheses pertaining to extramarital affairs revealed that a comparison of frequency responses to the question regarding incidences of extramarital affairs from the two groups, across the board, showed no significant differences. When comparing frequency responses among males and females regarding this question, however, significant differences at the .05 level of confidence were found. This coincides with other research findings. Neubeck (1969:37) wrote: "The grass has been greener on the other side regardless of where the turf is located, Africa or Europe, Asia or America, in cold climates and in the warm ones, among the white, the black, and the yellow races; there are extramarital relations in every continent and regardless of language, adultery persists in the face of powerful taboos and the most consistent religious dogma." Notions of extramarital involvement for Latin Americans are almost exclusively associated with the male. Regarding this, researchers, (Dominguez, 1966; Wagley, 1968; Keen, 1967) all have a tendency to associate their findings to socioeconomic factors, and strong double standard practices inherent to this type of culture. Approximately 50% of the males, and 28% of the females responded as having engaged in extramarital affairs during the marriage. This finding correlates with Kinsey (Hunt, 1974) which revealed that 50% of married men and 25% of married women commit adultery. The results obtained from the question regarding views about the effects of a self-incurred extramarital affair may have in the success of the marriage revealed no significant differences at the group level. When comparing frequency responses of males and females to this question, however, significant differences at the .01 level of confidence was found. Approximately 42% of the males and 82% of the females responded positively. On the question about views pertaining to the effect an extramarital affair on the spouse's part may have in the success of the marriage, no significant differences existed at the group level. Also, when comparing the males and females of the two groups regarding frequency responses to this question, no significant differences were found. Approximately 78% of the males and 82% of the females responded positively. Since the questions above dealt with double standard issues, the findings are supported by researchers such as Wagley (1968), who, in regards to Latin American double standards of sexual morals stated that these are manifested early in childhood and continue after marriage. From findings obtained in questions regarding extramarital affairs and double standard issues it is concluded that males from both groups show higher incidences of extramarital affairs and tendencies to practice double standards. The implication is that the reason for engaging in an extramarital affair, or

in participating in double standards, may not be directly the result of ethno-cultural factors, but rather the result of changes in current social trends and stratifications, as well as sexual freedom. Overall findings related to the hypotheses tested in the study confirm initial postulations that there are significant qualitative differences between the two groups, and in some instances between males and females of the two groups.

ADDITIONAL FINDINGS AND IMPLICATIONS: An overview of interpersonal variables related to prior marriages, socioeconomic status, and religion revealed some interesting outcomes which closely correlate with prior intermarriage research findings. Results of data show that most of the males and females in both groups had been previously married, were in the upper socioeconomic bracket, and married interfaith. This was especially noticeable among the intercultural group, where it was also noted that there were more Latin American males (18) married to Caucasian American females than were Latin American females (14) married to Caucasian men. These findings are supported by researchers such as Mittleback and Moore (1968), Simpson and Yinger (1953), Kuo and Hasson (1976), who postulated that those who marry across ethnic boundaries are more likely to be those who have been previously married, who marry across their religious lines, and who come mostly from a higher socioeconomic status. Also, that there are more males than females. Consistently throughout the study's findings and results in each of the areas examined...Self-Concept Congruency, Marital Adjustment and extramarital affairs...the results showed that the men ranked higher than the females. This imbalance implies that there may be a need for developing a counseling program which would best serve the mental health needs of intracultural couples evolving from specific socio-cultural implications, and psychological dynamics of marital adjustment. In light of this, the development of an instrument designed to measure intracultural marital adjustment could be a possibility, especially since current instruments used to measure marital adjustment have not been validated for use with minority groups.

RECOMMENDATIONS: In view of the study's findings and implications, the following recommendations for further research are made: 1. A replication of the study using a greater number of subjects in order to establish intra and inter group and individual comparisons, addressing the variables of self-concept congruency, marital adjustment and extramarital affairs in a more in-depth manner, to determine which dyads would be most affected. 2. A comparative study of direct effects of extramarital affairs and double standard practices, on the marital success of intracultural and intercultural males and females. 3. A longitudinal study to determine whether those who were considered successfully adjusted, and exhibited better self-concept congruence in their marriages (despite incidences of extramarital affairs), continued to evidence improved marital adjustment, and self-concept congruence after a period of time. 4. A longitudinal study to exclusively examine the males in these groups, to determine causes for involvement in extramarital affairs and whether extramarital affairs, and views regarding double standards prevail, have increased or decreased over a period of time.

Research Assistant, James R. Santianna, Superintendent
Mental Health USAF Regional Hospital

Social and Familial Role Perceptions as Correlates of Self-Concept and Locus of Control in Adolescents

Robert J. Drummond
Walter G. McIntire
University of Maine
Orono, Maine 04469

Abstract

The research described here was designed to investigate the relative contributions of selected familial and social role perceptions to self-concept and Locus of Control in a population of 117 high school sophomores. The students were dependents of military personnel (59%) and natives of northern Maine (41%). Familial and social role perceptions account for 61.8% of the variance in predicting self-concept and only 27.5% in the prediction of Locus of Control.

Self-concept and Locus of Control are two of the most widely investigated personality constructs in research with humans. Each offers potentially useful hypotheses and models of behavior for understanding the behavior of individuals or groups. The popularity of these constructs can be attributed to many reasons. While they are structurally complex, each offers intuitive "explanations" for observable behavior, easily administered instruments for quantification of the constructs are available, and the models are interpretable and utilitarian in clinical or applied settings. Much of the empirical research has focused on these variables as mediators or predictors of specific behaviors, feelings, or cognitive states. Less attention has been given to antecedents of their development.

Locus of Control is a construct postulated by Rotter (1966) in his social learning theory and operates both as a conceptual model of control in specific situations and as a general expectency model for diverse situations. Phares (1976) in his review of the research noted that the role of locus of control in personal adjustment and psychopathology continues to hold great promise for increasing our understanding of human behavior. He concluded that research studies have tended to show that internals are superior to externals in their efforts at coping and gaining a measure of control over their environment, a concept with obvious research and clinical implications.

Self-concept as a construct has its roots in the work of such theorists as James (1890), Mead (1934), Maslow (1954), Snygg & Combs (1949) and Rogers (1951). Shibutani (1961) in an examination of the development of self-concept, postulated that it is formed on the basis of two sets of sensory cues: internal conditions directly experienced by the person, and the consistent responses of others. Schools provide one significant setting for these cues. Dolan (1978) articulated the many influences of schools of self-development: the physical setting, type of instruction, level of ability, expectations of performance and perceived reactions of others. The family is the other obviously universal setting of the cues and is acknowledged as such by all the theorists noted above.

PRECEDING PAGE BLANK-NOT FILMED

The research reported here was designed to investigate the relative contribution of role perceptions of significant others and selected demographic variables to the prediction of self-concept and locus of control in high school students.

METHOD

The subjects for this study were all the 10th grade students from a rural northern Maine school. There were 62 males and 55 females. Sixty-nine of the students were dependents of personnel assigned to an air force base in the community. The balance were children of permanent residents of the area, primarily agricultural.

Three instruments were administered to all subjects. The first was a demographic data sheet, the second a semantic differential which included nineteen social, vocational, and familial role concepts. The concepts included were: My Father as, My Mother as; Successful, Kind, and a Disciplinarian; My Father as a Husband; My Mother as a Wife; Brothers and Sisters, and Myself. The social-vocational concepts presented were High School, Principal, Classmates, Guidance Counselors, Teacher, Career Military People, Farmers, and Discipline. Each concept was rated on ten bipolar adjective pairs from a set developed by Osgood, Suci and Tannenbaum (1957). Scores for each individual on each concept were computed by assigning values from one to seven (negative and positive) to each point between the adjective pairs and by summing these values for the ten adjective pairs under each concept. The third instrument utilized was the short form Nowicki-Strickland Personal Reaction Survey (1973) for grades 7-12. The scale is a 20-item pencil-paper test having Yes-No response mode. Split-half reliability for a grade 9-12 sample was reported to be .74.

Stepwise multiple regression was used with Locus of Control and Self-Concept as the dependent variables.

RESULTS

Predicting Locus of Control

A multiple R of .525 was computed for predicting Locus of Control with 24 of the independent variables entered accounting for 27.5% of the variance. The results of the stepwise regression for predicting Locus of Control are presented in Table 1. The Career Military People concept from the semantic differential was the first variable to be entered and contributed 11.5% of the variance. The correlation between Career Military People and Locus of Control was -.399. Teacher from the semantic differential was the second variable entered and contributed 4.1% additional variance. The correlation between Teacher and Locus of Control was -.344. Discipline was the third variable entered, correlated -.299 with Locus of Control and contributed 2.4% of the variance. Mother as a Wife loaded forth, and correlated -.267. The role figures Guidance Counselor, Classmates, and Main Office Secretary each contributed between one and two percent additional unique variance.

Predicting Self-Concept

A multiple R of .786 was computed for predicting the Self-Concept of the students, with 22 of the independent variables entered accounting for 61.8% of the variance. Mother as a Disciplinarian was the first variable to be entered and accounted for 27.8% of the variance. This variable

correlated .527 with Myself. High School was the second variable entered, added 10.5% variance and correlated .434 with Myself. Father as Kind was next and contributed 9.2% additional variance and correlated .515 with Myself. Guidance Counselor was fourth and provided 3.1% unique variance. Locus of Control contributed 2.5% while Police, Brother and Sister, and Mother as Kind each contributed between one and two percent additional variance.

Table 1
Results of Stepwise Regression for Predicting Locus of Control by Role Perceptions and Demographic Variables

Variable	Multiple R	R Square	RSQ Change	Simple R	F Final Step
Career Military People	.339	.114	.114	-.339	.005
Teacher	.395	.156	.041	-.334	2.068
Discipline	.424	.180	.024	-.229	.382
Mother as a Wife	.445	.198	.018	-.226	2.994*
Guidance Counselor	.457	.209	.010	-.320	.660
Classmates	.469	.220	.011	-.116	.465
Main Office Secretary	.479	.229	.009	-.320	.806
Military Dependent-Native	.485	.235	.005	.081	.894
Sex	.491	.241	.005	-.028	.504
Police	.497	.247	.006	-.303	.769
Extra-Curricular Activity	.503	.253	.006	-.071	.696
Brother and Sister	.507	.257	.003	-.134	.402
Myself	.510	.260	.002	-.119	.282
Father as Kind	.513	.263	.003	-.262	.501
Mother as Disciplinarian	.516	.266	.003	-.195	.798
Father as Disciplinarian	.518	.268	.001	-.118	.381
Father as a Husband	.519	.270	.001	-.140	.237
Assistant Principal	.521	.271	.001	-.320	.107
Farmers	.521	.272	.000	-.192	.144
High School	.523	.273	.001	-.282	.210
Mother as Kind	.524	.274	.001	-.199	.121
Father as Successful	.524	.275	.000	-.155	.066
Principal	.524	.275	.000	-.271	.038
Mother as Successful	.524	.275	.000	-.124	.015

*p .05

Table 2
Results of Stepwise Regression for Predicting Self-Concept
by Role Perceptions and Demographic Variables

Variable	Multiple R	R Square	RSQ Change	Simple R	F Final Step
Mother as Disciplinarian	.527	.278	.278	.527	1.817
High School	.619	.383	.105	.434	3.669*
Father as Kind	.689	.475	.092	.515	6.577**
Guidance Counselor	.711	.506	.031	.088	4.655**
Police	.726	.527	.020	.373	1.952
Mother as a Wife	.739	.547	.019	.465	3.594**
Brother and Sister	.763	.583	.010	.419	1.343
Mother as Kind	.771	.594	.011	.423	.632
Father as Successful	.773	.599	.004	.213	.739
Career Military People	.777	.604	.005	.307	.412
Discipline	.779	.607	.003	.333	.318
Teachers	.780	.609	.002	.293	.074
Extra-Curricular Activity	.781	.611	.001	.183	.263
Father as a Husband	.783	.613	.002	.450	.406
Principal	.784	.615	.001	.205	.182
Sex	.785	.616	.001	.079	.192
Classmates	.785	.617	.000	.390	.089
Mother as Successful	.785	.617	.000	.480	.084
Main Office Secretary	.786	.617	.000	.329	.033
Farmers	.786	.618	.000	.260	.040
Assistant Principal	.786	.618	.000	.116	.027

*p .05
**p .01

Conclusions

The correlation between self-concept and locus of control (r=.119) suggests that we are dealing with essentially independent psychological constructs. Locus of Control appears to be essentially unrelated to the familial and social role perception surveyed. It is interesting to note, however, that the little variance accounted for is by what could be described as discipline or authority figures (Career Military People, Teacher, and Discipline). The correlation between Locus of Control and these variables suggest that adolescents with an internal orientation perceive these models more positively.

The familial-social perceptions do predict significant variance when self-concept is the dependent variable. The first six variables entered provide an intrigueing set of observations. The early entrance of Mother as a Disciplinarian (1st) and Father as Kind (3rd) suggest that adolescents with high self-concepts do not perceive their parents as adhering to rigid traditional role performance expectation of mothers as expressive and fathers as instrumental. The important role of the school in personal growth is suggested by the fact that positive perceptions of High School and the traditional public school "humanist", the guidance counselor, are predictors of high self-concept.

The reader is cautioned that the data presented here is based on fewer than the desired number of subjects, subjects from a very specific geographic area, and from a school with a relatively unique mix of students (one-half from military families).

References

Dolan, L. The Development of Self-concept in the Elementary School: A Search for Determinants. Paper presented at the Self-concept symposium, Northeastern University, Boston, Sept. 1978.

James, W. Principles of Psychology. New York: Holt, 1890, 2 vols.

Maslow, A.H. Motivation and Personality. New York: Harper, 1954.

Mead, G.H. Mind, Self, and Society: From the Standpoint of a Social Behaviorist. Chicago: University of Chicago Press, 1934.

Nowicki, S. and Strickland, B. A Locus of Control Scale for Children. Journal of Consulting and Clinical Psychology, 1973, 40, 184-194.

Osgood, C.E., Suci, G.J. and Tannenbaum, P.H. The Measurement of Meaning. Urbana: University of Illinois Press, 1957.

Phares, E.J. Locus of Control on Personality. Morristown, New Jersey: General Learning Press, 1976.

Rogers, C.R. Client-Centered Therapy. Boston: Houghton Mifflin, 1951.

Rotter, J.B. Generalized expectancies for internal versus external control of reinforcement. Psychological Monographs: General and Applied, 1966, 80 (1, Whole No. 609).

Shibutani, T. Society and Personality. Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1961.

Snygg, D., and Combs, A.W. Individual Behavior: A New Frame of Reference for Psychology. New York: Harper, 1949.

Difference in Family Coalitions and Hierarchies Between Normals and Neurotics

Vance Shepperson
Mental Health Clinic (USAF Hospital)
Tyndall AFB, Florida

Abstract

This study compared the family coalition structures of 12 normal and 12 neurotic family triads matched for age, education, SES, and hospital exposure using a semiautomated behavioral role-playing task. Normal families were chosen from medical wards of a general hospital and reported no prior history of psychiatric treatment; neurotic families were chosen from psychiatric files in the same hospital and manifested: a) disturbed marital relationships, b) neurotic symptoms in at least one member (but no psychotic or characterological symptoms), a "constricted" emotional climate,and/or underachievement problems in the male adolescent. Results indicated that: a) the primary coalition within the normal group of families was between mother and father with a secondary father-son coalition also present; the primary coalition within disturbed families was also between mother and father but with a secondary coalition between mother and son, and, b) normal families appeared to manifest a relatively differentiated hierarchical family structure (father-mother-son) whereas disturbed families manifested a poorly differentiated, egalitarian family structure with a tendency toward maternal dominance (mother-father-son).

Family researchers, theoreticians, and therapists have for years been fascinated with what coalition and hierarchical differences might exist between well adjusted families and families experiencing difficulty coping adequately (e.g., Bandler, Grinder, and Satir, 1978; Haley, 1963, 1967, 1972; Minuchin, 1974). Within this interest area a body of research has suggested that a distinguishing hallmark of normal families is a strong primary coalition between father and mother with all other intra-family coalitions secondary to this one (Lennard, Beaulieu, and Embrey, 1968; Lidz and Fleck, 1960; Mishler and Waxler, 1968). On the other hand there is reason to suspect that in many disturbed families the primary coalition between mother and father is secondary to or sabotaged by strong competing coalitions between mother and son or daughter (Bowen, 1960; Lidz, 1963).

Another group of studies investigating family structure has asserted that normal families will tend to generate a well differentiated, hierarchical family pecking order in which father>mother>children (e.g., Mishler and Wasler, 1968). Families with "process" (chronic) schizophrenic or neurotic members will generate a poorly differentiated, egalitarian family structure in which control is more diffused (e.g., father=mother=children; Bowen, 1960; Leighton, Stollak, and Ferguson, 1971; Schuham, 1972).

Many legitimate methodological criticisms have been leveled at these studies and other related studies in this area (e.g., Jacobs, 1975, for an excellent review article). Much of this work relies heavily on client and therapist self-reports, case studies, surveys, or poorly defined and controlled empirical efforts with questionable validity and reliability. This current effort was designed to investigate the relative coalition and hier-

PRECEDING PAGE BLANK-NOT FILMED

archical structures within normal and neurotic families using a well controlled, semiautomated behavioral procedure with established reliability and validity. It is predicted that normal families will generate a stronger marital primary intra-family coalition on between group comparisons and a hierarchical (rather than egalitarian) family structure on within group comparisons; it is also predicted that disturbed families will generate egalitarian family structures on within group comparisons.

Methods

Personnel

Subjects. The subjects in this experiment were 12 disturbed and 12 normal family triads (mother, father, and adolescent son) matched for age, education, socio-economic status, and hospital exposure. Normals were taken from the records of admissions to a general (military) hospital medical ward and had no reported history of psychiatric treatment in the nuclear family; disturbed families were drawn from psychiatric records at the same hospital. This latter group of families had all indicated they were experiencing marital problems, neurotic symptoms in one or more family members (but no psychotic or characterological symptoms), evidenced a constricted family atmosphere in the judgment of their treating clinician, and/or presented underachievement problems in the adolescent son. This latter group of families was modeled after Riskin and Faunce's (1970) group of "moderately disturbed" families which was successfully separated from four other groups of families by a principle component factor analysis.

Experimenters and judges. The primary experimenter was responsible for selecting, contacting, and greeting the families upon their arrival as well as debriefing them after the experiment and paying them each $15 for participating. He was also responsible for training each of the judges to a minimum inter-rater reliability level ($r = .85$) for each of the videotaped behavioral dependent variables before the experiment proper. First year clinical psychology graduate students acted as judges on each videotape to gather relevant data; they were paid $3/hour for their work by the primary experimenter.

Materials. The audiotaped stimulus scenes used in this experiment were adapted to a family situation from those individual applications researched by Eisler, Hersen, and Miller (1975); Eisler, Miller, and Hersen (1973); McFall and Lillisand (1971); McFall and Marston (1970); and McFall and Twentyman (1973). Each scene called for a family member, systematically counterbalanced among families, to defend his or her rights from being violated by the remaining two family members or acquiesce to this violation of personal rights. The list of behavioral criteria used to evaluate the family interaction was grouped together to form an interaction Guide; this guide was adapted from Eisler, et al. (1973, 1975).

The dependent variables, frequency and duration of eye contact and number of verbal statements from each subject to another within the family triad was also taken in part from Eisler, et al's (1973, 1975) work.

These variables were also chosen on the basis of Alexander's (1970) recommendations that family process work be based on unobstrusive behavioral measures woven into the context of an intrinsically interesting task (e.g., family disputation) in as natural a setting as possible.

Procedure. Each family was selected from psychiatric or medical records

and sent a preliminary letter signed by the hospital administrator, then telephoned by the experimenter inviting them on a voluntary basis to participate. When each family arrived at the University of South Florida counseling center for their appointment they were led to an 8'x10' experimental room arranged such that three comfortable chairs, grouped around a small coffee table, were facing a one-way mirror. The experimenter mentioned that all necessary instructions were on the cassette audiotape recorder on the coffee table. He showed them how to work the recorder, asked if they had any questions, and said he would be back of the one-way mirror videotaping their discussion.

Instructions from the audiotape informed the family members that they were to be presented with six situations which might typically happen at home. They were instructed to role play their responses to the situation just as if it were happening right at that moment. The target to each stimulus scene (father, mother, son) was systematically counterbalanced to control for possible order effects. After a trial situation designed to insure that they understood the instructions, the six situations were presented. The following is a typical stimulus scene:

> "You knew it was going to be a long day. Your office was going to be inspected that day. Before you left for work in the morning you asked your son to do some yard work and you asked your wife to fix your favorite meal for supper. As you are driving in the driveway that night you notice the yard work is not done. You find TV dinners in the oven. Right then your wife and son walk in the kitchen carrying packages from their afternoon shopping spree. One of them notices you and asks how you are doing" (bell rings, which is the cue for target subject to begin his role playing).

After this role playing certain other procedures were implemented (Subjects taking a paper and pencil inventory and further role playing by each family member along with confererates matched for age and sex with missing family members.) These measures were designed to assess variables of assertion and aggression not relevant to this current discussion (see Shepperson, 1977).

Results

Differences in family coalition structure between groups.

Analysis of the verbal and non-verbal data suggest support for the predicted coalition structure within normal family (i.e., primary mother-father coalition with secondary father-son coalition). The data also suggests that the disturbed families' primary coalition is also between mother and father but with a secondary mother-son coalition (see Table I).

The data presented in Table I below is complicated by the fact that not only are the dependent variables significantly interrelated (mean correlation = .46) but the independent variables are interrelated as well; due to this blurring of variance a direct binomial probability estimate of the two dependent variables with least inter-correlation, frequency of eye contact and frequency of verbal units (r =.31), is the best approximate probability estimate: $6/\overline{5}(1/6)^2(1/6)^{2}\overline{/} = .0231$. From this calculation we can conclude that the differences in the above patterns between groups are statistically significant; that is, there is a meaningful difference in

the flow of verbal and non-verbal communication within family triads between diagnostic groups.

Table I. Means and standard deviations (SDs) of dependent variables in coalition data taken from the family context role playing sessions. The unit of analysis is the average sum of each family member's communication to the other (e.g., mother to father + father to mother = unit of analysis within father-mother dyad). N = 24 within each cell.

Variable	Dyadic Communication Pattern	Normals' Means and SDs (raw scores)		Disturbeds' Means and SDs (raw scores)	
Pattern of eye contact duration (seconds)	Father-mother	77.05	43.02	69.99	42.98
	Mother-son	44.87	29.37	38.35	21.41
	Father-son	62.33	48.49	29.32	13.71
Pattern of eye contact frequency	Father-mother	15.29	9.31	12.21	5.79
	Mother-son	11.25	5.19	11.04	5.43
	Father-son	14.08	5.31	8.79	3.47
Pattern of verbal unit frequency (number of sentences)	Father-mother	19.46	8.99	16.72	8.75
	Mother-son	9.08	5.80	9.04	6.47
	Father-son	13.75	5.70	6.62	4.07

Differences in family coalition structure within diagnostic groups.

A different analysis of the verbal and non-verbal data presented in Table I indicates support for the idea that normals would generate a hierarchical family structure and disturbeds an egalitarian family structure. The raw data was analyzed to determine which individual in a dyad looked longer or more frequently at another, or talked more frequently at another. The data for normals indicated that wives looked longer and more frequently at their husbands than vice versa (duration, $t = 3.18$, $p<.01$; frequency, $t = 2.79$, $p<.01$) while their husbands tended to talk longer than they did ($t = 3.40$, $p<.01$). Normal sons also looked longer and more frequently at their fathers than vice versa while they received more verbal units from their fathers (duration, $t<1$; frequency $t = 3.21$, $p<.01$; verbal, $t = 3.32$, p .01).

Within the disturbed family unit a less well defined interaction pattern was discovered. The only significant directional difference between spouses indicated that husbands looked longer at their wives and sons than vice versa (wives, $t = 4.06$, $p<.001$; sons, $t = 4.24$, $p<.001$). Disturbed fathers also looked more frequently towards sons and spoke more to their sons than vice versa (frequency, $t = 2.07$, p .05; verbal, $t = 2.28$ $p<.05$). Disturbed mothers, on the other hand, received longer and more frequent visual scanning from sons in addition to receiving more verbal units from them (duration, $t = 1.54$, $p<.1$; frequency, $t = 4.83$;

$p < .001$; verbal, $t = 2.13$; $p < .05$).

Discussion

Differences between groups.

An assumption of this research is that differential frequency and durations of verbal and non-verbal communication between individuals in a family group is indicative of relative coalitions within the group (e.g., longer and more frequent communication indicates a stronger coalition than does shorter frequency and duration of communication). If one grants this operational interpretation of a theoretical notion then we may summarize the obtained data by stating that the normals' parental coalition was stronger than that of the disturbeds' coalition. Further, the secondary coalition between groups differed, with normals evidencing a stronger father-son union than disturbeds and disturbeds evidencing a stronger mother-son coalition than normals. These results are consistent with the original hypothesis.

Differences within diagnostic groups.

An assumptive underpinning of the interpretation of the data within groups is that the most powerful person in a dyad or triad will speak the longest and receive the longest and most frequent visual scanning from others in that group.

Within the normals' diagnostic group the data support the idea of paternal dominance when interpreted according to this speculative framework. Wives and sons looked significantly longer and more frequently at father than father looked at them whereas fathers talked significantly longer to wives and sons (than they were spoken to by their wives and sons). No significant differences between mother and son were noted; however, since mothers are in a primary coalition with fathers this implies that the mothers' status within the family is more powerful than the sons' position. Thus, the normal family structure is fairly well differentiated in a hierarchical fashion with father>mother>son in terms of their relative status within the family triad.

Within the disturbed group of families one finds a less well defined interaction pattern (relative to the normals' pattern). The only significant directional difference in communication between dyads is that fathers looked longer toward wives and sons than vice versa, looked more frequently towards sons than vice versa, and spoke more to sons than vice versa. In contrast to this father-son pattern, mothers received longer and more frequent visual scanning from their sons in addition to receiving more verbal units from them than vice versa. Thus, the mother is the object of longer and more frequent visual contact and verbal contact from both fathers and sons than vice versa. When interpreted according to the speculative framework presented at the beginning of this discussion it would seem fair to conclude that neither a pattern of paternal or maternal dominance is present; further, these families' structures are less differentiated than normals and tend to manifest an egalitarian family structure with a slight tendency toward maternal dominance (mother≥father>son), as reflected in the non-verbal deference pattern which both father and son manifest toward her.

Conclusions

Two final comments, one regarding content and the other regarding

methodology, seem appropriate. First, constricted neurotic families would appear to profit from strengthening the parental subsystem while blocking development and maintenance of the mother-son subsystem; further, these rather passive disturbed fathers might well profit from verbal expressivity training in assertiveness. Second, we recommend the continued use of videotaped modalities within the family field (e.g., each judge needed an average of 19 passes on each videotape to gather relevant data). Researchers utilizing videotape modalities while focusing on unobstrusive dependent variables (e.g., eye contact) in an engaging, high interest task (e.g., family disputation) will maximize both the validity and reliability of the information they obtain.

References

Alexander, J. Videotaped family interaction. Paper presented at Western Psychological Association. Los Angeles: April, 1970.

Bandler, A., Grinder, J., and Satir, V. The Structure of Magic. Palo Alto, California: Science and Behavior Books, 1978.

Bowen, A. A family concept of schizophrenia. In D. Jackson (Ed.) The Etiology of Schizophrenia. New York: Basic Books, 1960.

Eisler, R., Hersen, M., Miller, P., & Blanchard, E. Situational determinants of assertive behavior. Journal of Consulting and Clinical Psychology, 1975, 43, 330-340.

Haley, J. Critical overview of present status of family interaction research. In J. Framo (Ed.) Family Interaction: A Dialogue Between Family Researchers and Family Therapists. New York: Springer, 1972.

Jacob, T. Interaction in disturbed and normal families. Psychological Bulletin, 1975, 82, 33-65.

Leighton, L., Stollak, G., & Ferguson, L. Patterns of Communication in normal and clinic families. Journal of Clinical and Consulting Psychology, 1971, 36, 252-256.

Lennard, H., Beaulieu, M. & Embrey, N. Interaction in families with a schizophrenic child. Archives of General Psychiatry, 1965. 12, 166-183.

Lidz, R., & Fleck, S. Schizophrenic patients and their siblings. Martal Psychiatry, 1963, 26, 1-18.

McFall, R., & Twentyman, C. Four experiments on the relative contributions of rehearsal, modeling, and coaching to assertive training. Journal of Abnormal Psychology, 1973, 81, 199-218.

Mishler, E., & Waxler, N. Interaction In Families: An Experimental Study of Family Processes and Schizophrenia. New York: Wiley, 1968.

Riskin, J. & Faunce, E. Family interaction scales, III: discussion of methodology and substantive findings. Archives of General Psychiatry, 1970, 22, 526-537.

Schuham, A. Power relations in emotionally disturbed and normal family triads. Journal of Abnormal Psychology, 1970, 75, 30-37.

Shepperson, V. The Behavioral and Self Report Assessment of Assertive Behavior, Aggressive Behavior, and Family Coalition Structure in Disturbed and Normal Family Triads, unpublished doctoral dissertation, University of South Florida, 1977.

John Henry Syndrome: Mans Conflict with Automation

Duncan L. Dieterly, Major, USAF
Chief, AFHRL Technology Office
NASA-Ames Research Center

Abstract

The purpose of this paper is to present some ideas and insights about the concept of automation and its impact upon the employee. An analogy is drawn using a folk hero to emphasize the extent of the problems encountered when dealing with the introduction of automation within an organization. The current state of automation research is briefly alluded to and some proposed directions for an improved understanding of its effects upon employee's efficiency are discussed.

INTRODUCTION

John Henry was a steel driving man. John Henry died with a hammer in his hand. This paraphrase of the closing lines of the song and tale that immortalizes the death of John Henry, a folk legend, dramaticly symbolizes the struggle of man against technology. In any event, his tale, which may be based upon the life of a real individual, emerged in the 19th Century as a tribute to his prowess, strength and integrity. To briefly recapitulate the legend, John Henry was a man of unusual physical and personal prowess. He worked his way around the United States: in the cotton fields as a picker, on the docks as a stevedore, on the railroad as a spike driver; always moving, always displaying a phenomenal physical superiority. He finally culminated his career as a steel driver on the construction of the Chesapeake and Ohio Railroad, in the rustic West Virginia hills, during 1873. The occupation of steel driver involved drilling the holes for the dynamite charges used in blasting the railroad tunnels. This was accomplished with the use of a sledge hammer and steel rod drill. John Henry, as always, excelled above all other men in his capacity to drive steel. He drove steel with a sixteen pound sledge hammer in each hand. It was during this time that the demon of technology in the form of a steam drill caught up with John Henry. A contest was arranged between John Henry and the steam drill. It was during this herculean battle that John Henry fell victim to the unrelenting, undefeatable machine. Although John Henry maintained a pace slightly ahead of the steam drill for over 10 hours, his body and spirit could not sustain the inevitable physical deterioration. John Henry died with a hammer in his hand.

Although this myth has many interpretations it accents man's concern with forces designed to constrain or restrain him. It is oddly paradoxical that mankind's own creativity embodied within technological advancement is viewed as an enemy to mankind. It is a struggle, perhaps, perceived by the combatants, as a symbolic contest between individual independence (physical strength) and societal dependence (using machines). The physical strength of man is the initial statement of independent action. The acceptance of technology then becomes an admission of weakness or acceptance of dependence, upon technology. Although technology was developed to enhance man's

capability it appears that not all men wish to utilize its advantages.

Automation is a word generally used to symbolize the introduction of new equipment to enhance man's capability to accomplish a required function. The steam drill mentioned in the story equalized men by providing the physical stamina most did not have. The introduction of the steam drill was perceived as a threat to individuals who had considerable prowess in this skill. Although the classical drama of automation is played against the backdrop of machines replacing mankind's physical skills we are now faced with the level of technology which may also replace mankind's intellectual skills.

The Federal Aviation Administration is currently studying the use of computers to conduct decision-making functions now performed by human controllers in an attempt to deal with the projected traffic growth. The program is known as AERA: Automated En-Route ATC. The technology would allow the computer to make decisions about conflict resolution, generation of clearances and their automatic transmission with the operator standing by to take over in an emergency. This system brings advantages and problems. The two major problems are how do the controllers retain proficiency and how to introduce the system so that both controllers and airspace users retain confidence in the ATC system (Klass: 1979).

The struggle is perceived as one between two groups not within one group. The crux of the problem is that the employee is not the one to introduce the new equipment and therefore has only partial knowledge of its advantages. The lack of knowledge festers into a feeling of loss of control because the individual is only thinking about themselves in terms of a fixed job. A job in this context is a set of known functions. If the job became dynamic or individuals became flexible enough to perceive that the changing of the job or the reduction of the number of jobs only means an improved opportunity to perform other functions, then automation may be perceived as a beneficial concept.

JOHN HENRY SYNDROME

The introduction of automated equipment generally results in the restructuring of the required interaction of the operator and the equipment. A function reallocation takes place, either formally as documented in operating instructions or informally as operators become familiar with the idiosyncracies of the new system. The spectre of automation in the organization generates a rash of symptoms which compose what I call the John Henry Syndrome.

The major symptoms of the John Henry Syndrome may be characterized within three areas: 1) quality enhancement 2) time reversal and 3) displacement jitters. Each of these symptoms is displayed to some degree when management begins to discuss the introduction or increase of automation within the organization.

In an international study on automation, Jacob and Jacob (1979) provide some interesting data on the answer to survey questions about three aspects of automation. The interesting finding is that workers in non-Communist countries do not perceive automation as improving their environment and are more fearful of displacement than in Communist countries. In general it is this type of attitude set which faces management before new equipment arrives.

The quality enhancement symptom is manifested by excessive discussions of the quality of the present product or service in terms of the individuals skills or capabilities. Emphasis upon the demand for craftsmanship and personal attention to insure a better product or service will be discussed. Stories to dramatize the effect of unusual excellence or substandard quality will be recited with glee.

The time reversal symptom will be evident by increased concern with how things were before and how the world is going to the dogs. The wish to return to simpler, basic, safer times will become apparent. Much of the talk will center upon how previous managements respected and placated their employees.

The displacement jitters are characterized by allocating more time to discussing retirement, other job opportunities and general dreams of some level of independent status. This symptom may also cause increased friction between employees and an increase in flaring tempers may be noted. A surge of individual motivation may be experienced, briefly, as the employees try one more time to justify their existence before falling prey to the complete fear of what tomorrow will bring. The symptoms are readily identifiable as subtle changes occur in the work climate. If management will indicate the development of the John Henry Syndrome and should alert them to the appropriate action. In the example mentioned earlier of the AERA system a plan of transition was developed to preclude a John Henry Syndrome.

GENERAL MODEL

In order to understand the impact of automation a general systems model was developed. The model attempts to switch the focus from the machine system to human resource and machine interface. The model draws upon the classic man-machine interface concept and emphasizes the operator functions and the machine functions. It also allows for a set of functions which may be accomplished by either the operator or the machine. The immediate environment is the location of the system operation. The procedural aspects are the state methods of operations either passed along by word of mouth or documented in a 'handbook for operations.' The more complex the system the higher the probability of procedural directions including considerations beyond the operations of the system.

Change in the function allocation and the resulting system performance that occur from introducing automated systems may impact the operator performance, immediate environment or the total system effectiveness. The key issue is that the introduction of automated equipment displaces the previous procedural system and redistributes functions. These changes introduced in a system may have the potential to impact the entire organization depending upon its degree of dependence upon the system. Therefore as the organizational set of functions, considered to be a series of job sub-sets, is increasingly allocated to the machine(s), the higher the dependence of organizational effectiveness upon the automated system. A change that specifically modifies the operator's functions may have a cascading impact on the total human resource or the organization.

DIMENSIONS OF INTEREST

Although the concept of automation has not been directly studied as a variable of interest, it is a dimension of systems design that is critical to the operator function, maintenance function and procedural function. The maintenance function is assumed to be located in the immediate environment of the system. Although the maintenance structure of a specific organization may be far more complex, for our purposes it will be considered located in the immediate environment. Maintenance functions are those required to maintain the equipment. The degree to which maintenance is dependent upon automated equipment introduces another interactive set of job functions. In order to understand the immediate impact of automation then five dimensions are given prominence: operator functions, equipment functions, maintenance functions, procedural guides and system performance.

The constant improvements in technology are producing weapon systems of extreme complexity. The current expansion within the computer industry spearheaded by IBM, which believes it may have a prototype three pound computer capable of 70 million instructions per second (MIP's), a fourteenfold increase over today's systems, is an indication of the rapid growth in technology to be considered in designing new systems. One factor forcing the complexity is the increase in use of automated systems to perform sub-functions within the total operational sphere. The availability of automated systems and the subsequent perceived improvement in reliability is indisputable. The effect of adding to or enhancing a system with automated subsystems is always anticipated as an improvement in total system performance. However, the battlefields are strewn with systems that were not successful due to some aspect of equipment failure, operator failure or organizational failure.

IMPORTANCE

The ability to understand and predict the impact of automation in terms of human resources requirements in the future may save an organization considerable resources. The impact of automated systems being introduced into advanced weapon systems, commercial aircraft and all areas of industry could be a determining factor in the success or failure of the organization. As automated systems increase in capability and cost the critical trade off of cost for effectiveness cannot be avoided. The adaptations of a massive word processing system, command and control systems or automated plants may require a billion dollar investment. Therefore, a precise understanding of what type of system performance may be attained in terms of the dimensions cited is critical.

When an automatic system is introduced into a system, the first visible impact occurs when the system becomes operational and the automatic system must be implemented by the operator. Since all systems have undergone rigorous testing and evaluation by skilled test specialists and engineers, no problems are anticipated. However, the normal operational environment is more variable and demanding than the test environment and the operators are performing at a different experience level than the specialists. Therefore, when the system becomes operational if there are any difficulties associated with the new automatic sub-system, they will first appear through errors in the operator-machine interface.

The next level of impact will be experienced through the procedure-

operator interface. A new system, which may be employed without procedures or established practices being considered nor designed will produce a series of system failures. Frequently systems are designed in terms of machine functions but they do not consider the operator functions nor the resulting mix. This is especially crucial for emergency operations. (Dieterly: 1979)

The new automatic system will also impact maintenance functions which, in turn, increase maintenance training requirements. The basic engineering concerns of reliability and maintainability are only the first level of possible machine difficulty. Changing procedures or introduction of new procedures may cause considerable upheaval in the operator population and/or maintenance population. Although throughout the entire system development phases efforts are made to reduce these types of problems, there are currently no adequate data sources or guides that may be applied.

FUNCTION ALLOCATION

The complexity of functions and excessive work load experienced during critical system phases demands a different approach. The necessary functions to be accomplished per operational segment must be considered as the basic element. The allocation of these functions between man and machines must then be determined. The current system has an established function distribution pattern, new systems do not. When an automatic system is introduced, a clear indication of change in operator function must be understood and considered. Automation should be considered in terms of function allocation. That is, how do the function distributions within the system change due to the introduction of automation. Emphasis should be placed on the functions allocated to the human resources in the system. The assumption has to be made that the functions allocated to the machine have attained a level of reliability to generally preclude failure. A separate, an important area of concern is the degree to which the operator is willing to make this assumption in the operational environment. If the operators are not convinced that the system is reliable, then they will adapt a strategy which establishes informal cross checking behavior in addition to the prescribed procedural behavior. This type of strategy will increase workload, perhaps beyond the operator's capability.

To automate a part or a total system usually results in three outcomes to the existing system functions: (1) new functions are introduced, (2) previous functions are deleted, and (3) existing functions are changed. The introduction of new equipment, such as a HUD (Head up display) for tactical weapon delivery, may appear to be the ideal way to increase system performance. However, if the HUD, in addition to all other instruments, does not provide the necessary information the operator demands, it will increase workload because attention will be split between the HUD and other information sources. If the operators perceive the HUD unreliable, then they will not depend upon it. In either case, the expected system performance will not be attained.

If a HUD is introduced with all instruments removed that it replaces, then the operator may lose confidence in the aircraft and employ heuristic methods for accomplishing the task. This type of situation will also result in poor system performance. What we are seeing is the beginning of a

John Henry Syndrome which must be conteracted to preclude failure. If the HUD is introduced and procedures are changed, then the operator is required to retrain to a level that insures he will not revert back to the former procedure. Again, anticipated improvements in system performance will not occur. The same problem will be magnified as the impact on maintenance and other organizational elements is addressed.

Unfortunately , the human factors analysis and the engineers analysis are not inter-related. Each accomplished its objectives but only a system analysis of the man and machine in terms of function allocation can tell the full story. As was indicated earlier, research concerned with the second condition (deletion of functions), especially to the level of deleting total jobs, predominate studies documented in industry. Much of what occurs in the Air Force falls under the first condition (adding functions). In this case, more automated equipment is introduced and the operators, procedures and maintenance functions are increased. An understanding of the interaction of the operator, procedure and maintenance functions and the function distribution as contributory aspects of system performance is required before the introduction of automation can be studied.

A major difficulty in studying automation is to find an agreed upon workable definition. Rather than delve into the subtle arguments for different approaches , I would like to suggest that there is no absolute why of defining automation and that individuals perceptions of system performance and the operator functions determine the level of automation. If this is accepted assumption then a given system as understood or perceived by different individuals may have different degrees of automation. An aircraft equiped with an automatic pilot may be perceived as automated by the manufacturer but not by the operator who does not use the system except under extreme emergencies. Although, this idea complicates the issues it must be considered to understand the resulting system performance.

CONCLUSION

Automation is one aspect of technical advancement. It must be viewed as a variable to be considered in determining total system performance. The introduction of automation changes the function allocation between the operator and the equipment. The operator's perception of the equipment in terms of their individual skill is critical. With the advent of computer technology many of mankind's most burdensome tasks can be delegated to the machine. Man can therefore anticipate more time to be a creative, thinking, individual. Before mankind can reap the benefits of automation, the early warning symptoms of the John Henry Syndrome must be identified and alleviated. Whatever the medicine, it must be administered. Heroic battles may still be fought, but martyrdom to automation is no longer appropriate. Do not die with a hammer in your hand.

Bibliography

- Encyclopedia Britanica, Vol 13: USA: 1966 (p 39)

Dieterly, D.L. "Decision/Problem Analysis in Accident Reconstruction." Journal of Safety Research, Winter 1979.

Jacob, B.M. and Jacob, P.E. Automation and Humanization, draft of Research Corporation of the University of Hawaii Report, 20 Aug 79.

Klass, P.J. "Decision Computers Studied for ATC." Aviation Week & Space Technology, Nov 79, 37-39.

Pugh, C. "New Eyes for the Giant," Airman, Dec 79, 30-33.

Behavioral Contributions to Strategic Planning

V.P. Luchsinger, Brig, Gen. USAFR
Texas Tech University, Lubbock, Texas 79409
Otis Bryan, Lt. Col. USAF
HQ, USAF (SAY), Washington, D.C. 20330

Abstract

The behavioral sciences are viewed for their inclusion in strategic (long-range) planning for the defense establishment. Implications for implementation of behavioral inputs into planning decisions are surveyed for relevance and potential contribution to the effectiveness of planning.

With an increasingly complex world, the search for tools and insights to contend with the uncertainty of the future increases yearly. The higher stakes involved in the impact of Department of Denfense ventures and activities make mistakes more costly and success more important. Science and methodological fields of inquiry are being consulted for guidance into the quest for better decisions concerning the future.

Strategic planning is a vital tool of the large organization in the private sector when decisions about the future are involved. Strategic planning in the civilian sector is also referred to as corporate planning or long-range planning. The essential idea remains the same. A view is taken of the world of the future to help improve the futurity of present decision. Presumably, better outcomes should follow better planning. Major organizational decisions always have momentous future implications. To preclude a totally military semantic sense of strategic planning, it should be stated that strategic planning is used in the sense of executive level deliberations, long range implications, and a broad organizational scope of decisions affecting goals, objectives, programs, and policies of the organization.

Inputs into organizational strategic planning have been dominated by economic, geopolitics, and management sciences. There has been a concerted effort to conceptualize, organize, and analyze data and situations with which military planning has been concerned. The U.S. Air Force has recognized the importance of long-range planning by establishing a Deputy Directorate for Long Range Planning in the DCS, Operatic. s, Plans, and Readiness staff unit.

The recognition of the importance of long range planning in the Department of Defense is accompanied by the increased demands placed upon intellectual disciplines to deal with the futurity of notions such as national interest, power, inflation, and resources. A reflection of behavioral science influence is seen in the utility of such constructs and approaches as decision theory, communication networks, systems analysis, group processes, and the like.

Testimony for the inclusion of such behavioral contributions has ranged from Welch speaking at the Military Operations Research Society meeting (1979) to Brzezinski (1969). There has been a call for the incorporation of behavioral sciences into policy research at the national level

to facilitate successful implementation of programs (Goodwin, 1971). Horowitz (1967) has called for increased contributions by social and behavioral scientists into policy and programs, as well as the traditional research role.

Methodologies have varied. In his classic piece, Etzioni (1969) suggested some extension of individual correlates of behavior to the national and international scene. For example, he extended psychological defenses mechanisms such as denial, emotional isolation, and habituation to national and international behaviors. We see these behaviors today.

The level of analysis and generalization may vary. The Etzioni minimalist approach argues that nations are different from persons. His maximalist view claims that the behavioral study of man has no limitation in its contribution to the affairs of man. Another view claims that war is a social institution. As a traditional outlet for collective human conflict, was should be capable of more acceptable solutions for dealing with conflict than the traditional (but technologically accelerated) forms of violence.

Hamsher and Sigall (1973) have provided eloquent substantiation for behavioral inputs into policy sciences on the national and international level. This paper would propose a conceptual framework for the application of behavioral sciences to strategic (albeit long-range or corporate) planning for the Department of Defense. Inputs of a behavioral nature could be categorized as internal and external to the organization, with external being further characterized as domestic and foreign. If behavioral science can assist the organization to cope with its various environments, a "goodness of fit" needs to be observed. That is where behavioral sciences can provide fine tuning for the applications of other sciences and judgement to policy and strategic planning. The ultimate case for the behavioral sciences rests on the fact that organizational affairs depend upon people making decisions about what people in organizations will do to cope with environments that must be contended with.

The internal environment of the organization for strategic planning consists of micro issues that affect the operational behavioral of the organization as a whole, and participants and units in particular. What we know about the behavioral sciences can be related as follows:

1. Motivation is a continuing challenge in organizations. Strategic planning cannot assume motivation after the fashion of economic resources. Issues such as recruiting, retention, productivity, and work environment are rife with implications of motivation for planning. Yet, planning seldom seems to deal with motivation, but hopes it won't become a problem.

2. Learning behavior is an important construct for the coping mechanisms of the organization. Training, readiness, career development deal with the applications of learning theory. This is an area where the armed forces have a huge investment and do an excellent job operationally. Yet the contemporary stress on rapid development forces in strategic planning requires an adaptation of behavior to help support the appropriate response to a threat to national interest. Adaptive coping behavior is the essence of learning, which should be reflected in strategic planning.

3. Decision theory runs through the fabric of present defense strategic planning. Behavioral inputs could help affirm the impact of

goal and objectives-setting on performance, as well as recognize aspects of subjective probabilities and human uncertainties that defy the assumptions of perfect rationality. After all, we don't always deal with rational, optimizing adversaries.

4. Organization theory has much to offer strategic planning. Concerns over boundary spanning, conflict, uncertainty absorption, and contingency theory alone could make significant contributions to organization design and effectiveness.

5. Human resource development is a major area of behavioral contribution. Ackoff (1970) cites frequently the case of corporate planning which slights human planning as a crucial resource. Issues of personnel acquisition, retention, development, retirement, minorities, morale, and assignments seriously affect the people segment of the defense budget. Human asset accounting cries out for behavioral applications to economic processes.

6. Change is a fundamental topic of study in the behavioral sciences. Strategic planning is concerned with the design and execution of organizational and national response. Yet we fail to practice what behavioral scientist preach about change and conflict.

In the external context of the organization, we are sensitized to issues in the environment which must be accepted for, and with which we must cope. Considering the domestic environment first, some behavioral variables include:

1. Attitudes. As predispositions for behavior, domestic attitudes are of crucial importance. Attitudes seriously affect defense strategic planning. Attitudes toward the military vary with the times, but they don't just happen. McCloskey has presented an intriguing hypothesis that individually and nationally it can be seen that attitudes are a function of personality . The concept of a national "personality" alone is provocative.

2. Demography. The quantitative and qualitative nature of population change is a serious factor in national strategy. It affects the armed forces today and in the future. It is about to affect other national powers such as the Soviet Union. The nature of the supply of the human resource has importance that is difficult to overestimate.

3. Social values. The value systems of society have much to do with the kind of defense a society thinks it wants and gets. McClosky (1960) presented some hypotheses that suggested that behavior such as isolationism was associated with extreme fears, resentments, and personal insecurities. Aggressive tendencies in national and international behavior were seen as consequences of external and internal frustrations. The social psychologist and sociologist should have much to provide of value to planning that is sensitive to the impact of values. Progressive marketing firms already practice such analysis in their strategic planning.

In the case of foreign or international environments, values, attitudes, and demography also are relevant. However, other factors that involve behavior would include:

1. The concept of risk. If a risk function can be behaviorally derived for a person, can this be done for a country? Why and when does the Soviet Union become more venturesome in an area of the world involved with

multiple risk? Bozeman (1976) cites the proper consideration of deterrence strategies to depend upon the perception of risk. This risk is a behavioral variable as much as a political and economic construct.

2. Stress is found throughout the environment. It is particularly evident in the international context. Kelman (1965) claimed that the psycho-mental roots of war were entwined with the tolerance for stress and conflict. Certainly, defense strategic planning should consider behavioral aspects of stress as well as other considerations. History is rife with behavioral impact on the conduct of war.

3. Intercultural and comparative studies would contribute from the material available of cross-cultural findings. Etzioni and Christiansen (1959) placed great emphasis on foreign policy which was a resultant of attitudes. Yet understanding of differences were seen as tension reducing mechanisms between states and peoples.

All in all, this is a beginning statement for the case of behavioral science contributions to strategic planning. Certainly, behavioral implications are evident in the observed success of failures of planning. It often appears that behavioral aspects of planning are done poorly, or not at all in many realms of long range planning. For the consideration of the variable nature of behavior, it would seem important to do at least some contingency planning that includes behavioral inputs. The contemporary planning practice of "environmental scanning" certainly could include some behavioral assessments of variables important to the planning involved. After all, people make plans for people to execute. What better justification for the study of behavior as the proper prerequisite for making the decisions about the future of organizations.

In summary, the case for behavioral input into defense planning might be borrowed from the preamble to the Constitution of UNESCO as follows:

"War begin in the minds of men"

Perhaps the minds of men can make contributions to helping man end war.

References

Ackoff, R.L. A Concept of Corporate Planning. New York: Wiley, 1970, 154p.

Bozeman, A.B. "War and the Clash of Ideas." Orbis. 1976. 20(1) 61-102.

Brzezinski, Z. "Purpose and Planning in Foreign Policy. Public Interest. 1969, 14, 52-73.

Christiansen, B. Attitudes Toward Foreign Affairs as a Function of Personality. Oslo: Oslo University Press, 1959, 282pp.

Etzioni, A. "Social-Psychological Aspects of International Relations." Lindzey and Aronson (Eds.). Handbook of Social Psychology. Vol. V, Reading, MA: Addison-Wesley, 1969, 538-601.

Goodwin, L. On Making Social Research Relevant to Public Policy and National Problem Solving." American Psychology. 1971, (V) 431-442.

Hamsher, J.H. and Sigal, H. (Eds.) Psychology and Social Issues. New York: MacMillan, 1973, 548pp.

Horowitz, I.W. Social Science and Public Policy. International Studies

Quarterly. 1967, 1, 32-62.
Kelman, H.C. International Behavior: A Social-Psychological Analysis. New York: Holt Rinehart Winston, 1965.
McCloskey, H. Perspectives on Personality and Foreign Policy. World Politics. 1960, 13, 129-139.
Welch, J.A. Applications of Behavioral Sciences. Presented at Military Operations Research Society, West Point. June 1979.

The Risks of the Cuban Missile Crisis: Controlling the Analysts

Richard Ned Lebow
The National War College
Washington, D.C. 20319

Of all the decisionmaking groups dealt with by students of international crisis, Kennedy's Ex Com has received the most extensive praise from theorists. Roger Hilsman, Alexander George, and Irving Janis, to name just three, are lavish in their encomiums for Kennedy and to a great extent base their suggestions for improving the quality of crisis decisionmaking on Kennedy's management of Cuba.[1] Something of an aura has in fact come to surround those 13 days in October. They are exalted as a glorious exercise in rational policymaking that should be emulated by other leaders. Just how realistic is this portrayal?

Students of the Cuban crisis point to the various ways in which Kennedy attempted to encourage free-wheeling debate about the various policy options open to the United States: he charged the members of the Ex Com with considering the problem of the Soviet missiles in Cuba as a whole rather than confining themselves to their specific areas of authority or expertise; he purposely included officials in the Ex Com whom he knew were likely to represent differing points of view; he called in outside experts with this end in mind; he encouraged the group to debate the pros and cons of several different options without committing himself to any of them in advance. The President also absented himself from some of the Ex Com's sessions in the hope of encouraging more frankness on the part of his advisors. In his absence either Robert Kennedy or Ted Sorensen was assigned the role of "intellectual watchdog" in order to ensure that every policy matter was analyzed thoroughly. During some of these sessions the participants engaged in role playing, acting as advocates for positions they did not necessarily support. At one point the Ex Com divided into two subgroups, each assigned with reaching a policy decision and defending it before the other.[2]

Kennedy's efforts were unquestionably innovative and succeeded in bringing about a more thorough evaluation of the various policy options than might otherwise have been the case. It would nevertheless be an exaggeration to describe the Ex Com as an open decisionmaking environment. The Ex Com's mandate was a narrow one: to consider the pros and cons of the variety of coercive measures that could be employed to get the Soviet missiles out of Cuba before they became operational. From the very outset, the President made it clear to the group that he would neither acquiesce to the presence of the missile nor seek to remove them by purely diplomatic means. He was intent on using force to

PRECEDING PAGE BLANK-NOT FILMED

overcome the threat posed by the missiles and charged the Ex Com with recommending what military option was best suited to the task. In effect, he made the most important policy decision before the Ex Com even convened. The group never debated the wisdom of using force, despite the realization by all of the participants that such a course of action risked triggering a nuclear war. Even Janis, one of Kennedy's greatest admirers, admits that "the Executive Committee could be criticized for conforming too readily with the President's way of defining its mission."[3]

But Kennedy's formulation of the problem did meet with some opposition at first, McNamara believed that Soviet missiles in Cuba made no real difference to the strategic balance. At most, he argued, it permitted the Soviets to close the missile gap in 1962 instead of a few years later when their second generation ICBM's came on line. According to Hilsman: "The clear implication of McNamara's position was that the United States should do nothing, but simply ignore the presence of Soviet missiles in Cuba and sit tight."[4] Adlai Stevenson was also initially opposed to using force to get the missiles out. On October 16, Kennedy told the ambassador that Soviet missiles had been detected in Cuba, and voiced his conviction that they would have to be taken out by an airstrike. Stevenson was shocked by the President's apparent determination to resort so readily to violence and wrote him a note urging caution. In the note, which he personally delivered the following morning, Stevenson warned:

> To risk starting a nuclear war is bound to be divisive at best, and the judgments of history seldom coincide with the tempers of the moment. . . . I feel you should have made it clear that the existence of nuclear bases anywhere is negotiable before we start anything. . . . I confess I have many misgivings about the proposed course of action.[5]

Kennedy was quite annoyed by the note. He brushed aside Stevenson's objections as well as McNamara's, by specifically excluding from the agenda of the Ex Com the possibility of a diplomatic approach to Moscow. By most accounts McNamara was bludgeoned into accepting the need for forceful action. Along with his deputy, Roswell Gilpatric, he adopted the blockade, which held out the prospect of the least overt use of force, as a fallback position. Stevenson also came out in support of the blockade but continued to voice his concern that everything possible be done to avert war. On October 21, when the National Security Council considered what diplomatic action might accompany the blockade, he raised the possibility of striking a deal with the Russians. Stevenson proposed that the United States give up its base at Guantanamo and guarantee the territorial integrity of Cuba in return for the demilitarization and neutralization of that country. As an alternative, he urged the

group to consider offering to withdraw American Jupiter bases in Turkey and Italy as a quid pro quo for a Russian withdrawal of their missiles from Cuba. United Nation's inspection teams would subsequently ensure that none of the foreign bases of either superpower were used to mount a surprise attack.[6]

Stevenson's proposals were made in the context of the blockade and would not, he insisted, seem "soft" if they were properly worded. But the President rejected them out of hand.[7] Stevenson was also subjected to a sharp attack by members of the Ex Com led by Lovett and McCone. Allison cautions against drawing too many inferences from this exchange but there seems little doubt that Stevenson was ostracized by the core of the Ex Com. The President's cavalier treatment of him probably encouraged other members of the group to give vent to their emotions. Allison himself speculates that Kennedy "may have sacrificed the Ambassador to the hawks in order to allow himself to choose the moderate, golden mean."[8] Whatever the explanation, Stevenson, who had been asked by the President to return from New York specifically for this meeting, was deeply wounded by the gratuitously vindictive nature of the attack upon him. According to Abel, "The bitter aftermath of that Saturday afternoon stayed with him until his death."[9]

If Stevenson questioned the overall strategy of the Ex Com, its members had been at loggerheads all week over tactics. Neither advocates of the blockade nor those of the airstrike were able to bring about a consensus. The Ex Com's ultimate decision in favor of the blockade was the result of the strong Presidential pressure. Sorenson reports that "The President was impatient and discouraged" by the fourth day of the Ex Com's deliberations. "He was counting on the Attorney General and me, he said to pull the group together quickly--otherwise more delays and dissension would plague whatever decision he took."[10] Kennedy made it clear that he wanted to act by Sunday and that to do so he needed a decision in favor of the blockade. When this failed to materialize at the next meeting of the Ex Com, Sorenson invoked the President's authority in order to achieve a consensus. He announced "that we are not serving the President well, and that my recently healed ulcer didn't like it much either."[11] The group got the message and the following day rallied to the blockade. On Saturday, the decision about what response to make to Khrushchev's second cable was brought about in the same way.[12]

The reality of the Ex Com does not measure up to the myth propagated by Kennedy's admirers. The Ex Com proved a relatively pliant tool of the President. Knowledge of his preferences shaped its deliberations at every turn as none of the participants were prepared to speak out in favor of a position they knew the President would not support. Even when Kennedy did not attend the group's meetings the prospect of a

free-wheeling debate was inhibited by the presence of his brother and Ted Sorenson, whom everyone expected would report what was said back to the Oval Office.[13] Independent thinking was tolerated only within the limits set by the President. Officials who expressed unacceptable points of view were pressured like McNamara to bring their opinions into line or were personally abused, as happened to Stevenson.

The attack on Stevenson, one of whose proposals ironically became the basis for the resolution of the crisis, appears to be a classic manifestation of groupthink. The Ex Com had emerged from 5 days of intensive deliberations with a remarkable degree of group solidarity but a somewhat fragile policy consensus. It was prepared to defend the blockade option before a wider circle of officials; the National Security Council meeting on Saturday actually marked the group's debut in this respect. Stevenson was an outsider. His proposals challenged the consensus and by extension the solidarity of the Ex Com. This solidarity was unquestionably important to many Ex Com members as a means of coping with the extraordinary stress of a nuclear crisis. Their otherwise uncalled for attack on the mild mannered Stevenson is best interpreted as a mechanism by which their sense of solidarity could be expressed and strengthened.

In the final analysis the Ex Com could be described as a superb example of promotional leadership. It was brought into being less to make policy than to legitimate it. Kennedy's choice of its members, its restricted agenda and the use of Robert Kennedy and Ted Sorenson as policemen all point to this intent to use the Ex Com as a means of building a consensus for whatever specific course of action he ultimately decided upon. In practice, the deliberations of the group influenced policy by helping to shift Kennedy away from the airstrike in favor of the blockade although Allison suggests that the Ex Com was only one of several influences in this direction.[14] Kennedy was preparing to initiate a confrontation that he knew risked war with the Soviet Union. To be effective he needed widespread bi-partisan support. If something went wrong and the two superpowers moved toward an even more serious confrontation, he knew that he would need this support even more. Kennedy's adroit if not fully conscious manipulation of group dynamics helped to create that political backing. Allowing the Ex Com to debate the pros and cons of the major action-oriented options encouraged them to believe that they were instrumental in making policy, as indeed to a certain extent they were. The group solidarity that developed in the course of these proceedings helped to transform otherwise disparate individuals with different political outlooks and bureaucratic loyalties into staunch supporters of the blockade and the other initiatives which accompanied it. The united front they presented impressed other governmental officials and congressmen and helped to widen the scope of support for the President.

The evidence of promotional leadership and groupthink in the Cuban case raises important doubts in this author's mind about the extent to which leaders are willing and able to take steps to overcome these kinds of decisionmaking pathologies. As we have noted, Kennedy's handling of Cuba has become a template against which several distinguished decision-making theorists believe the performance of other crisis managers ought to be measured. If Kennedy was only partially willing to permit an open decisionmaking environment, what can be expected of other leaders in other situations?

NOTE: An expanded version of this paper will appear as part of Chapter VIII of Between Peace and War: The Nature of International Conflict. (Baltimore: The Johns Hopkins University Press, 1980).

1. Roger Hilsman, To Move A Nation. (Garden City, New York: Doubleday and Co., 1967), pp. 159-232; Alexander George, "The Case for Multiple Advocacy in Making Foreign Policy," American Political Science Review, LXVI (September 1972), pp. 751-95; Irving Janis, Victims of Groupthink. (Boston: Houthton, Mifflin, 1972), pp. 138-66.

2. Abel, supra; Allison, supra; Hilsman, pp. 159-232; Janis, pp. 138-66; Sorenson, pp. 667-718.

3. Janis, p. 143. Janis appears to turn a blind eye to practically all of the instances of groupthink and promotional leadership in the Cuban case. The reader cannot help but be struck by Janis's apparent need to portray the Cuban case in a good light perhaps in order to have a case that proves it is possible to make decisions in crisis without succumbing to groupthink.

4. Hilsman, p. 195.

5. Sorenson, pp. 694-95.

6. Ibid., pp. 695-96.

7. Ibid.

8. Allison, p. 209.

9. Abel, p. 96.

10. Sorenson, p. 692.

11. Ibid.

12. Allison, p. 227.

13. Ibid., p. 207.

14. Ibid., p. 202.

THE VARIABLES THAT PREDICT USER SATISFACTION WITH RESEARCH: A MULTIPLE REGRESSION APPROACH

William C. A. Stewart and Thomas J. O'Connor
Applications and Liaison Office, AFHRL
Brooks AFB, Texas 78235

Abstract

The Air Force Human Resources Laboratory produces over 100 research projects annually. This paper analyzes the extent of user satisfaction with the research. A questionnaire was sent to 54 users of research and a stepwise multiple regression algorithm was used to explore several hypotheses concerning the relative importance of seven predictor variables in determining overall satisfaction with the research. The results indicate that user satisfaction depends upon a small number of variables. The best predictor of user satisfaction was the clarity of recommendations variable [$r = .67$, $F(1,52) = 41.87$]. It can be concluded that research has a higher probability for use if the researcher stresses the identified predictor variables for user satisfaction in the research report.

Introduction

The Air Force Human Resources Laboratory (AFHRL) conducts the Air Force's largest personnel research program, with a budget of over 18 million dollars for fiscal year 1980. A critical issue for AFHRL, and indeed, for any such user-oriented research organization, is the degree of user satisfaction with its research. As one part of its mission, the Applications and Liaison Office of AFHRL addresses this issue directly by querying the user of each of its research projects about the degree to which the research satisfied stated requirements.

However, even if the degree of user satisfaction with a research project is known, more penetrating questions remain. For example, one may ask what aspects of the research the user considers in determining his degree of overall satisfaction. If there are identifiable aspects of a research project that determine the user's overall satisfaction, and if there is consistency among users in the relative importance they attach to these various aspects, then it should be possible to construct a weighting system such that, given the degree of user satisfaction with each of the several aspects, the degree of overall user satisfaction could be predicted. Such a system, were it successful, would constitute an analysis of the variable, Overall Satisfaction, into its components, and would, in addition, reveal the relative importance of the components as determinants of Overall Satisfaction. This paper describes a project used to construct such a system.

Method: A Multiple Regression Approach

By means of a comprehensive questionnaire, the Research Utilization Evaluation Report, each requestor of an AFHRL research effort was queried concerning a number of aspects or variables of the research. Two areas covered in the Evaluation Report will be the focus of the paper. The first area identifies seven variables of the research and asks the user to indicate, on a scale of from 1-5, the degree of satisfaction with each of those variables. The format is as follows:

How satisfied are you with each of the following aspects of this research: (Fill in the blanks with one of the numbers below)

1. Very satisfied
2. Somewhat satisfied
3. Neither satisfied nor dissatisfied.
4. Somewhat dissatisfied.
5. Very dissatisfied

_____Timeliness of response
_____Completeness of study
_____Clarity of recommendations
_____Relevance of findings to the problem
_____Feasibility of implementation
_____Projected cost of implementation
_____Projected benefits of implementation

The second area of the Evaluation Report concentrates on overall user satisfaction with the research. The user was asked to indicate degree of overall satisfaction with the research, using the following format:

Did this research satisfy the user needs? (Circle one of the responses.)

1. Very satisfied
2. Somewhat satisfied
3. Neither satisfied nor dissatisfied
4. Somewhat dissatisfied
5. Very dissatisfied

The set of predictor variables are then used in an algorithm in an attempt to predict the criterion variable, Overall Satisfaction.

The first fifty-four research projects completed by AFHRL in CY 1978 were evaluated by the users of the research and their Evaluation Reports constitute the data for the paper.

Results

The seven predictor variables and their abbreviations are listed in Table 1. The intercorrelations of the predictors are given in Table 2, and Table 3 presents the correlations of each of the predictors with Overall Satisfaction. From Table 3, it can be seen that the Clarity of Recommendations variable produced the highest correlation with Overall Satisfaction, and an analysis of variance revealed a highly significant effect of linear regression, $F(1,52) = 41.87$, $p < .001$.

Using all seven predictors to predict Overall Satisfaction resulted in a multiple R of .717, indicating that over 50% of the variation in Overall Satisfaction can be attributed to variation in the predictors. An analysis of variance revealed a highly significant effect of linear regression, $F(7,46) = 6.97$, $p < .001$, supporting the conclusion that several of the major determinants of Overall Satisfaction have been identified.

However, an examination of Table 3 reveals several fairly substantial correlations between predictors. For example, the correlation between COMPLETENESS and CLARITY is .827, and the correlation between RELEVANCE and CLARITY is .808. Thus, some of the predictors may be redundant, in the sense that removing them would not significantly reduce the proportion of

criterion variance accounted for. In the present case, a reduction in the number of predictors would not only lead to a more economical prediction system, but would also identify those predictors that are most important for prediction of overall satisfaction.

A stepwise regression algorithm was used to determine which predictor variables should be retained in the prediction system. A detailed description of the algorithm is given by Efroymson (1962), and Draper and Smith (1967), and only a brief outline of the procedure will be given. Variables are entered into the prediction system one at a time, and a variable is allowed to enter the system only if its inclusion results in a significant decrease in the error mean square. In addition, at each step of the procedure, each variable already in the system is examined to determine if it would have led to a significant decrease in the error mean square, had it been the most recently entered variable. Thus, a variable which had entered the system at an earlier step may be removed at a later stage because of its relationships to other variables.

Table 1

Predictor Variable Abbreviations

Timeliness of response	TIMELINESS
Completeness of study	COMPLETENESS
Clarity of recommendations	CLARITY
Relevance of findings to the problem	RELEVANCE
Feasibility of implementation	FEASIBILITY
Projected cost of implementation	COST
Projected benefits of implementation	BENEFITS

Table 2

Intercorrelations of the Seven Predictors of Overall Satisfaction

	TIME-LINESS	COMPLETE-NESS	CLARITY	RELE-VANCE	FEASI-BILITY	COST	BENEFITS
TIMELINESS	1.000						
COMPLETENESS	.517	1.000					
CLARITY	.553	.827	1.000				
RELEVANCE	.501	.697	.808	1.000			
FEASIBILITY	.358	.654	.763	.779	1.000		
COST	.463	.588	.676	.668	.741	1.000	
BENEFITS	.465	.698	.729	.734	.733	.628	1.000

Table 3

Correlation of the Seven Predictors with Overall Satisfaction

TIMELINESS	.494
COMPLETENESS	.617
CLARITY	.668
RELEVANCE	.621
FEASIBILITY	.542
COST	.401
BENEFITS	.524

Application of the stepwise algorithm revealed that the variable BENEFITS could be omitted from the prediction system without leading to a significant decrement in predictability. The value of multiple R for the reduced prediction system was .717, indicating virtually no loss in predictability. The prediction equation specified by the reduced system is as follows:

Overall Satisfaction = .211(TIMELINESS) + .169(COMPLETENESS) + .326(CLARITY) + .225 (RELEVANCE) + .149 (FEASIBILITY) - .266 (COST)

COST enters the equation with a negative regression weight. Since the simple correlation between Overall Satisfaction and COST is positive, it appears that COST is acting as a suppressor variable in the prediction system. One possible explanation of the effect of the COST variable is that a user's satisfaction with other aspects of a research project is reduced by a tendency to believe that significant research projects will tend to be costly. If so, then subtracting the proportion of variance due to COST will improve predictability of overall satisfaction.

Conclusion

Researchers should be aware of the components that contribute to user satisfaction with research, lest the research go unused. With various programs, including research, competing for funds, it is imperative that research be presented to the user in a manner that maximizes its potential use.

References

Draper, N.R., and Smith, H. Applied Regression Analysis, John Wiley & Sons, Inc., New York, 1966.

Efroymson, M. A., Multiple regression analysis. In Ralston, A. and H. S. Wolf (Eds), Mathematical Methods for Digital Computers. John Wiley & Sons, New York, 1962.

Antihistamine Radiation Protection is Dissimilar From Estrogen Radiation Protection[1]

G. Andrew Mickley
Armed Forces Radiobiology Research Institute
Bethesda, Maryland 20014

and

Department of Behavioral Sciences and Leadership
United States Air Force Academy, Colorado 80840[2]

Rats suffer an early transient performance decrement immediately after a sufficiently large dose of ionizing radiation. However, it has been shown that males experience a more severe incapacitation than do females. This sex difference has been attributed to the low estrogen levels in the male. In support of this notion, supplemental estrogens in castrated male rats have produced enhanced performances postirradiation. Antihistamines have also previously been shown to alleviate radiation's effect on behavior. The present study revealed that antihistamines are only effective in altering the behavioral incapacitation of sexually intact subjects. This contrasts with the previous work which indicates that estrogens can only benefit gonadectomized rats. These findings suggest different mechanisms underlying antihistamine and estrogren radiation protection.

Rats that have been exposed to a sufficiently large and rapid dose of ionizing radiation exhibit an early transient incapacitation (ETI) which is characterized by akinesia (Kimeldorf and Hunt, 1965) and decrements in motivated (avoidance) behaviors (Casarett and Comar, 1973). Typically, performance on a behavioral task is compromised for a period of up to 0.5 hr after the irradiation. This behavioral decrement is followed by temporary recovery, progressively poorer performance, permanent incapacitation, and death. This syndrome has also been noted in pigs (Chaput and Wise, 1970) and monkeys (Young, Chapman, Barnes, Brown, and Hurst, 1968) trained in a shock-avoidance task.

Although the ETI syndrome was identified years ago, its exact etiology is still quite uncertain. One specific hypothesis is that the accumulation of the biogenic amine histamine may contribute to the ETI. In this regard, Doyle and Strike (1975) demonstrated an increase in blood histamine after 4000 rads of ionizing radiation. This increase peaked at about 3 min postirradiation, decayed with time, and approached control values at 20 min (about when the subject's behavior returned to pre-irradiation baseline). Blood pressure is one physiological factor that is sensitive to histamine changes, and several investigators (Turbyfill, Roudon, and Kieffer, 1972; Bruner, Bogo, and Henderson, 1975) have demonstrated that severe peripheral hypotension often accompanies periods of postirradiation performance decrement. Pretreatment with the anti-

1. Research was conducted under AFRRI Work Unit Number MJ60112
2. Present address

histamine chlorpheniramine maleate effectively alleviates the hypotension associated with ETI and may improve performance (Turns, Doyle, and Curran, 1971).

Other experiments, from this laboratory, have shown that male and female rats differ in their response to ionizing radiation (Mickley, 1980). Females suffer less severe performance decrements after a large dose of radiation and recover more quickly than do males. This behavioral sex difference can be eliminated by gonadectomizing the rats and can be reinstituted by administering estrogens to these subjects.[3] Thus, estrogens, in addition to antihistamines, have been shown to play an important part in the relief of postirradiation performance decrements.

There are known interactions between histamine and gonadal hormones (Henningsson and Rosengren, 1972; Snyder and Axelrod, 1965) which suggests that these two systems could be reacting to radiation insult in similar ways in order to produce behavioral deficits. Alternatively, manipulation of either of these systems might be used to improve postirradiation performance.

The present experiment sought to determine if the mechanism behind antihistamine radioprotection might be similar in some ways to that of estrogen radioprotection. This was approached by determining if irradiated rats might be benefited by antihistamine pretreatment in a way that is similar to that already demonstrated by estrogen supplements (Mickley, 1980).

Methods

Subjects. Nineteen male and sixteen female experimentally naive Sprague-Dawley rats (200-300 g) were used in this experiment. In addition, thirteen male and twelve female rats of the same strain were gonadectomized under ether anesthesia within the first 5 days of birth, and then allowed to mature. All rats were housed three to a cage in a room illuminated from 7:30 a.m. to 4:00 p.m. daily. Ambient temperature was set at 20°C. Purina rat chow and water were continuously available.

Apparatus. All rats were trained to avoid foot shock by leaping a distance of 10 cm from the floor of a cage up onto a retractable lucite ledge. The avoidance apparatus consisted of a Plexiglas chamber 37x24x37 cm in height with a grid floor that could be electrified with scrambled shock. A signal click preceded the foot shock by 5 sec. The lucite ledge was moved into the cage at the onset of the click and retracted after the foot chock (0.5 ma for 10 sec). A microswitch on the ledge indicated if the subject successfully moved up to the shelf before the shock. The intertrial interval was held constant at 30 sec.

Procedure. All subjects were trained to meet an avoidance-response criterion of 100% over 50 consecutive trials. Rats received at least 50 trials once a day until this criterion was met. This usually took about 5 days. Ten minutes before irradiation 7 castrated males, 6 ovariectomized females, 13 intact males and 10 intact females were injected with 20 mg/kg chlorpheniramine maleate. i.p. The remainder of the subjects

3. Interestingly, intact males are minimumly benefited by estrogen treatments.

received saline control injections. The subjects were then irradiated, immediately after which 50 trials were again administered in the avoidance chamber.

The Armed Forces Radiobiology Research Institute (AFRRI) Linear Accelerator was used to expose rats to high-energy electron radiation. Each animal was irradiated individually with its right side to the source while in a constraining polyethylene tube. Electrons were accelerated to an energy of 18.1 MeV at a peak beam current of 0.44 amps. Pulse duration was 4 μsec, and pulses were delivered at a rate of 15/sec. In order to make the field more uniform, a water-scatter device was used which lowered the energy of the beam by 3.5 MeV. Each rat was positioned 3.5 m from the source and received a midline tissue dose of 10,000 rads (13 rads/pulse), which has been demonstrated to consistently produce an ETI (Mickley and Teitelbaum, 1978). Total exposure time was between 0.8 and 1.2 min. Dosimetry was accomplished by using 0.05-cc tissue-equivalent ion chambers whose calibration is traceable to the National Bureau of Standards. Animals were retrieved within 1 min after exposure. Previously published data indicate that constraint and sham irradation do not alter performance on this avoidance task (Mickley and Teitelbaum, 1980).

Results

Percent avoidance scores were computed for each animal during 8, 2.5 minute time periods: 1.0-2.5, 2.5-5.0, 5.0-7.5, 7.5-10.0, 10.0-12.5, 12.5-15.0, 15.0-17.5, and 17.5-20.0 minutes postirradiation (see figures 1 and 2).

During the first 2.5 minutes after irradiation all groups showed a significant ($p < .05$, sign test) decrease in avoidance responding as compared with the 100% baseline performance.

Avoidance scores were then used in the computation of an analysis of variance. The remainder of the present report concentrates on the drug variable i.e., comparisons of antihistamine effects with those of control injections. Comparisons were made during three postirradiation time periods (0-2.5, 5.0-7.5, and 15.0-17.5 minutes) which were representative of early, middle, and late ETI effects. There were significant drug effects observed during the early and middle time periods (0-2.5 minutes: $F=4.056$, $df = 1$, $p = .049$; 5.0-7.5 minutes: $F=11.588$, $df = 1$, $p = .001$). A significant drug X surgery interaction was found during the 15.0-17.5 minute time period ($F = 10.784$, $df = 1$, $p = .002$). Subsequent t tests revealed that, at all time periods surveyed, antihistamine pretreatment significantly improved the intact male performance over and above the control-injected males (0-2.5 min, $p = .017$; 5.0-7.5 min, $p = .00072$; 15.0-17.5 min, $p = 6.4 \times 10^{-10}$). The improvement was so dramatic that males treated with antihistamines were also significantly better than the intact females (0-2.5 min, $p = .00123$; 5.0-7.5 min., $p = .0075$; 15.0-17.5 min, $p = .05$). There was no significant difference between control-injected female rats and females injected with antihistamines (all $p > .05$).

Figure 2 illustrates the finding that gonadectomized rats were <u>not</u> helped by antihistamine injections. Ovariectomized females injected with antihistamines were not significantly different from control-injected ovariectomized females at any of the three time periods submitted to statistical analysis (t tests; all $p > .05$). Likewise, castrated males which

received chlorpheniramine did not differ from controls during either the 0-2.5 or 5.0-7.5 min time periods postirradiation (t tests; both $p > .05$). During the 15.0-17.5 minute postirradiation time frame, castrated males with antihistamine did significantly worse than controls (t test $p = 4.0 \times 10^{-11}$). Needless to say, castrated males treated with antihistamine were no different or significantly worse than intact females at all analyzed time periods postirradiation (t tests; 0-2.5 min, $p = .01$; 5.0-7.5 min, $p = .49$; 15.0-17.5 min, $p = 8.0 \times 10^{-11}$). While ovariectomized females with antihistamines performed more poorly than intact females, this difference was not significant at anytime postirradiation (t tests, $p > .05$).

Discussion

As reported elsewhere (Mickley, 1980), after irradiation control female rats perform better than males (see Figure 1) and gonadectomy eliminates this sex difference (see Figure 2). However, unlike the previous estrogen findings which indicated that estradiol was only effective in enhancing the performance of gonadectomized animals and ineffective in intact animals, the present results suggest that antihistamines produce behavioral benefits in intact animals but not gonadectomized ones. Thus, although estrogen and antihistamine both allow the irradiated animals which are benefited by these injections to perform at similar enhanced levels, the mechanisms behind the actions of the two drugs is apparently dissimilar. Although the gonadal hormone systems are known to interact in some fashion with the endogenous histamines (Henningsson and Rosengren, 1972; Snyder and Axelrod, 1965), this interaction may indeed be a complex one in the irradiated animal.

Acknowledgements: The author wishes to thank both Dr. Herman Teitelbaum, for his advice during the course of this project, and Mr. Bernard Dennison, for his fine technical support. The statistical advice offered by Dr. Dickie Harris is also gratefully acknowledged.

References

Bruner, A., Bogo, V., and Henderson, E.A. Dose-rate effects of ^{60}Co irradiation on performance and physiology in monkeys. Topical Report DNA-3660T, Defense Nuclear Agency, Washington D.C. 1975.

Casarett, A.P. and Comar, C.L. Incapacitation and performance decrement in rats following split doses of fission spectrum radiation. Radiation Research, 1973, 53, 455-461.

Chaput, R.L. and Wise, D. Miniature pig incapacitation and performance decrement after mix gamma-neutron irradiation. Aerospace Medicine, 1970, 41, 290-293.

Doyle, T.F. and Strike, Radiation released histamine in the rhesus monkey as modified by mast cell depletion and antihistamine. AFRRI Scientific Report 75-18, Armed Forces Radiobiology Research Institute, Bethesda, MD, 1975.

Henningsson, S.S.G. and Rosengren, E. Alterations of histamine metabolism after injections of sex hormones in mice. British Journal of Pharmacology, 1972, 44, 517-526.

Kimeldorf, D.J. and Hunt, E.L. Ionizing Radiation: Neural Function and Behavior. New York: Academic Press, 1965.

Mickley, G.A. and Teitelbaum, H. Persistance of lateral hypothalamic-mediated behaviors after a supralethal dose of ionizing radiation. Aviation, Space, and Environmental, Medicine, 1978, 49, 868-873.

Mickley, G.A. Behavioral and physiological changes produced by a supra-lethal dose of ionizing radiation: Evidence for hormone-influenced sex differences in the rat. Radiation Research, 1980, 81, 48-75.

Snyder, S.H. and Axelrod, J. Sex differences and hormonal control of histamine methyltransferase activity. Biochim. Biophys. Acta, 1965, 111, 416-421.

Turbyfill, C.L., Roudon, R.M. and Kieffer, V.A. Behavior and physiology of the monkey (Macaca Mulatta) following 2500 rads of pulsed mixed gamma-neutron radiation. Aerospace Medicine, 1972, 41-45.

Turns, J.E., Doyle, T.F., and Curran, C.R. Norepinephrine effects on early postirradiation performance decrement in the monkey. AFRRI Scientific Report 71-16, Armed Forces Radiobiology Research Institute, Bethesda, MD, 1971.

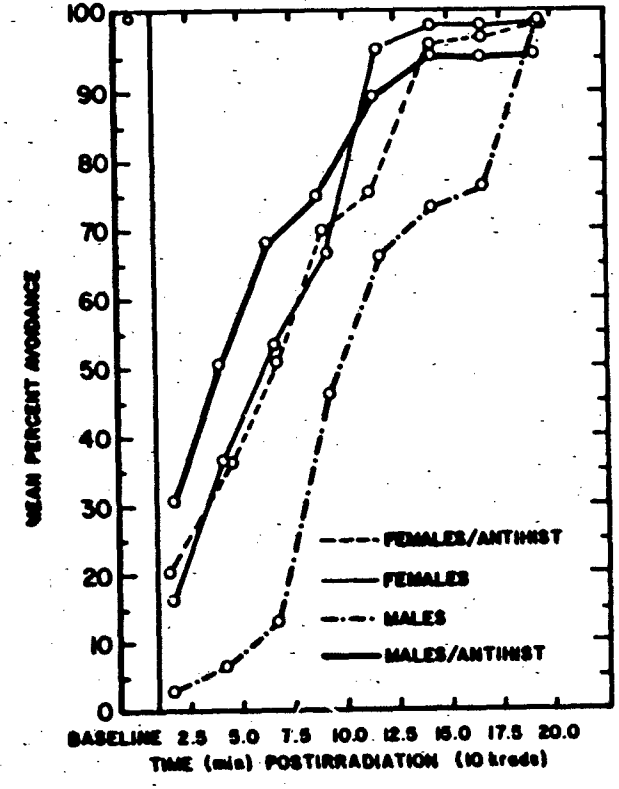


Figure 1. Postirradiation avoidance performance of sexually intact rats. Antihistamine decreases the severity of behavioral incapacitation.

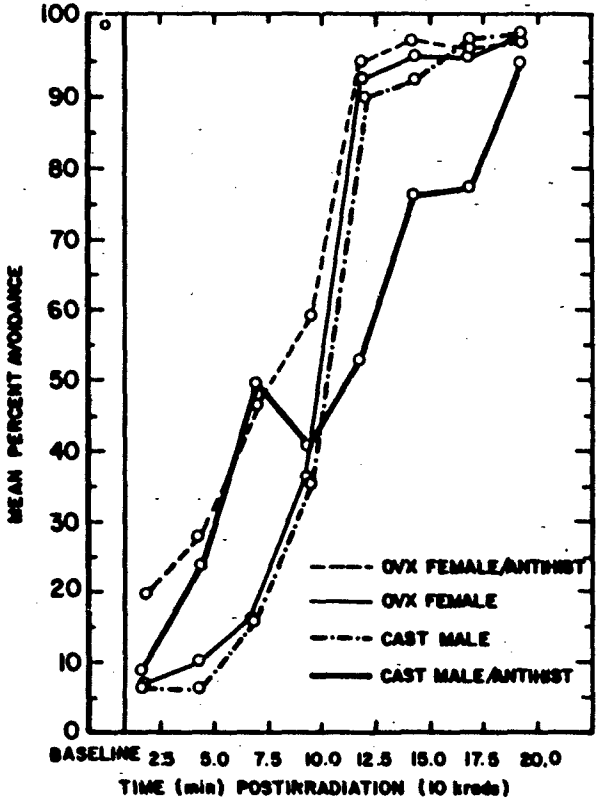


Figure 2, Postirradiation avoidance performance of gonadectomized rats. Antihistamine fails to enhance performance of neutered subjects.

Aircrew Combat Training Simulation: Perspectives and Views

Ronald G. Hughes
Operational Training Division
Air Force Human Resources Laboratory
Williams Air Force Base, Arizona 85224

Note: A complete transcript of this presentation is available upon request from AFHRL/OTR (Dr. Hughes), Williams AFB, AZ 85224.

Abstract

Aircrew combat training simulation is discussed as one means of overcoming skill deficiencies created by the inability to practice critical combat skills in a peacetime training environment. Research, development, and training efforts being conducted by the Operational Training Division of AFHRL are described in both the areas of air combat maneuvering and surface attack training. Special attention is given to work involving the A-10 and F-16 configurations of the Advanced Simulator for Pilot Training. A successful demonstration is described whereby F-16 and F-4 flight simulators located at different geographical locations were flown interactively . . . a significant first! Particular emphasis is paid to the development of a hostile threat environment for the A-10 in the surface attack role and to the training potential for such simulation.

In connection with reference to Air Force Systems Command's Project 2360, ideas are presented as to the full range of requirements for tactical aircrew training simulation. Ideas are presented under the general headings of (1) visual simulation of tactical terrain, (2) visual simulation of maneuver elements, (3) simulation of special visual effects, (4) weapons systems simulation, (5) simulation of command, control, and communication, and (6) instructional features for tactical training simulation.

The presentation concludes with a discussion of how such a simulation capability might be integrated into an overall training system. The notion of a geographically dispersed simulator network for broad scale combined arms training simulation is introduced but is dismissed in favor of a more general "generic combat simulator system." Such a generic system is discussed as having advantages not only for flight training applications but for applications to ground combat vehicle simulation as well.

Conclusion

Serious resource constraints dictate that new technologies be developed to cope with the need for maintaining a continued level of combat readiness. These technologies will result in a facelift to operational training over the next two decades that will be unparalled in history. This paper seeks to anticipate these new directions with special emphasis upon applications in the area of aircrew combat training simulation.

TACTICAL FLIGHT SIMULATION

by

Dirk C. Prather
Robert S. Kellogg
Carl H. Castore
Operational Training Division
Air Force Human Resources Laboratory
Williams Air Force Base, Arizona

ABSTRACT

In March 1979 the Operational Training Division of the Air Force Human Resources Laboratory conducted a feasibility study in tactical scenarios. Combat ready A-10 pilots flew a mission in which they were to destroy an enemy tank with their 30mm cannon. A modified air defense system including anti-aircraft and surface-to-air missiles was employed to thwart the A-10 pilot's mission. Each pilot flew twenty runs in which he ingressed to the target area, attempted to destroy the tank, and egressed out of the hostile environment. The learning curves indicate data similar to those of actual combat. The learning curves also indicate that training of this type should be beneficial in lowering loss rates in the first few missions a pilot would fly in combat. Research studies indicate the highest loss rate among pilots is from the first ten missions of combat and the Israeli's high losses during the first few days of attempting to penetrate a hostile defense system necessitates research be conducted in the area of tactics. The simulation of a NATO or Korean scenario, allowing man to fly and practice the tactics developed, has a high potential payoff for the defense posture of the United States.

INTRODUCTION

The purpose of this study was to test the feasibility of using the Advanced Simulator for Pilot Training (ASPT) in training pilots for combat in a simulated hostile environment. It has been suggested [1] that in the future, complex flight simulators will not be limited simply to procedures and instrument training, but will serve the much more important and broader need of training combat ready pilots in the complex task of tactics development. NASA's experience in successfully training crews to perform moon landings, which could only be practiced through simulation, has demonstrated the value of simulating tasks which are impractical or impossible to practice in the real situation. Experiences of the Air Force and Navy during the Viet Nam era indicated that pilot performance in a combat situation could be improved considerably by providing the pilot with experience in flying mock air-to-air engagements against dissimilar aircraft and against the types of ground defenses he would be coming up against in

PRECEDING PAGE BLANK-NOT FILMED

combat[2]. This type of experience provided the pilot with the opportunity to develop his skills at offensive and defensive strategies. The success of this program led directly to the establishment of the Red Flag and Blue Flag exercises by Tactical Air Command (TAC) and similar training exercises by other commands. While such exercises are excellent test beds for new tactics on a squadron or wing level, they are frequently of too short a duration for the evolution of new tactical approaches or an exploration of the relative effectiveness of alternative strategies on an individual pilot level. With the use of a high fidelity full mission simulator such as the ASPT, it is possible to expose the pilot to a large number of combat simulations of sufficient fidelity in a short period of time to provide him an optimum situation for offensive and defensive skill development. Using such simulation will engender a more holistic view of the combat mission in which the pilot will be required to meet the everchanging mission requirements with flexibility and judgment, based on "experience." The present study is a first effort to demonstrate that the simulator can be used in this extended role.

COMBAT ENVIRONMENT

Existing visual environments on the ASPT were used and modified to produce the computer generated imagery (CGI) used in this study. The imagery was presented in an A-10 simulator which included the heads up display (HUD). A map of the basic environment was given to each pilot. It consists of an area of about 10 miles square. Twelve mountains are included in the area, ranging in height from 1400 to 6500 feet. Also, there were a number of small rock-like projections between the mountain ranges on the right. Three surface-to-air missile (SAM) sites were located in the central valley. The SAM site closest to the aircraft was the one activated. If the aircraft was outside of its firing envelope, the other two SAM sites searched to see if the aircraft was within their range. When the aircraft was within range, a 3.5 seconds alternating 2025 Hz spaced tone indicated radar acquisition. Immediately after the acquisition warning, a three second alternating 2025 Hz tone indicated imminent launch. At the end of the tone, the SAM fired and a light illuminated to indicate launch. The missile could be seen by the pilot if he was in a position of proper viewing angle. A rapidly alternating tone continued for as long as the SAM was in flight. If a SAM had been launched and the aircraft subsequently exited the firing envelope of the site, the SAM would abort. If, during the time the SAM was in flight, the aircraft entered the firing envelope of another SAM site, the SAM in flight would continue its course unless the aircraft exited the envelope of the launching site, in which case the SAM aborted. If the aircraft was then being tracked by a new site, a new acquisition cycle was initiated. If hit by the SAM, the aircraft went out of control and crashed. Collocated at each SAM site was an anti-aircraft artillery battery (AAA), consisting of four 23 mm guns as well as an isolated AAA site to the west of the battle area. The AAA site began tracking when the aircraft flew within range. If the pilot took evasive action, which consisted of +3 G's or more, or -1.5 G's or less, for six consecutive seconds, the aircraft would evade AAA kill; if not, the

aircraft would be hit and go out of control. Muzzle flashes were the only indication that the aircraft was being fired upon. The aircraft did not have the capacity of destroying the SAM or AAA batteries. The enemy target in this environment was a tank which was located at one of six randomly selected positions on the road stretching between the two mountain ranges at the Southern end of the battle zone. The tank, once placed in one of the six positions, remained stationary. It showed muzzle flashes from its cannon, but did not have offensive capabilities. The A-10 aircraft had only the capability of firing its gun. The tank was considered destroyed when it was hit by one round during the first sequence of firing, in the two-dimensional space occupied by the tank. In addition, a special feature was added to the environment allowing the pilot to see his own shadow when his altitude was below 500' MSL.

Subject Population

The subjects for this study were combat ready A-10 pilots, as determined by Tactical Air Command Headquarters. They were all in Mission Qualification Training at Davis-Monthan AFB, AZ. They ranged in age from 26 to 35 years. Three of them had combat experience and all but two of them had participated in Red Flag exercises.

Experimental Runs

The experimental sequence for the subjects was as follows. Each subject was briefed individually by a research instructor pilot (IP) on the combat environment, including all the offensive and defensive characteristics. He was given a map of the environment which he was allowed to study before entering the simulator. He was allowed to take the map along during the simulator run. The pilot was initialized approximately ten miles north the forward edge of the battle area (FEBA) in IFR conditions at an altitude of 5500 feet. The pilot's only instructions were that he was on his own and that his mission was to ingress to the battle area, destroy the tank and egress safely across the FEBA. The experimental run was terminated when one of the following occurred: a. SAM kill; b. AAA kill; c. Terrain crash; d. Over G; and e. Safely crossed FEBA. As soon as each run terminated, the pilot was immediately reinitialized and another run began. Each subject flew a set of runs in the morning and a second set in the afternoon. Each set lasted between an hour to one hour and a half. Each subject was given feedback as to what happened at the termination point of each run: if he hit the tank, if he survived, and if he was shot down, the source of his demise. Between periods, the subjects were not allowed to discuss tactics among themselves so as not to bias their individual approaches to solving the tactical problems.

ANALYSIS

On the basis of whether the pilot hit or missed the target, survived, or was destroyed, a chi square analysis was performed. The overall analysis with all data pooled showed statistical significance at the 0.05 level.

This means that the pilot's offensive scores (hitting the tank) and defensive procedures (survival) are highly correlated. When morning performance was compared with afternoon performance, there was a significant increase in survival and offensive capabilities in the afternoon. All seven subjects showed increased capacity to perform offensively and defensively after repeated training runs in the simulator. The composite curve, shown in Figure 1, displays a clear acquisition phase at the beginning followed by a leveling off during the mid-trials and then a marked end spurt during the last few trials. The curve clearly shows that the pilots have learned how to perform better offensive and defensive maneuvers after repeated trials in the simulator.

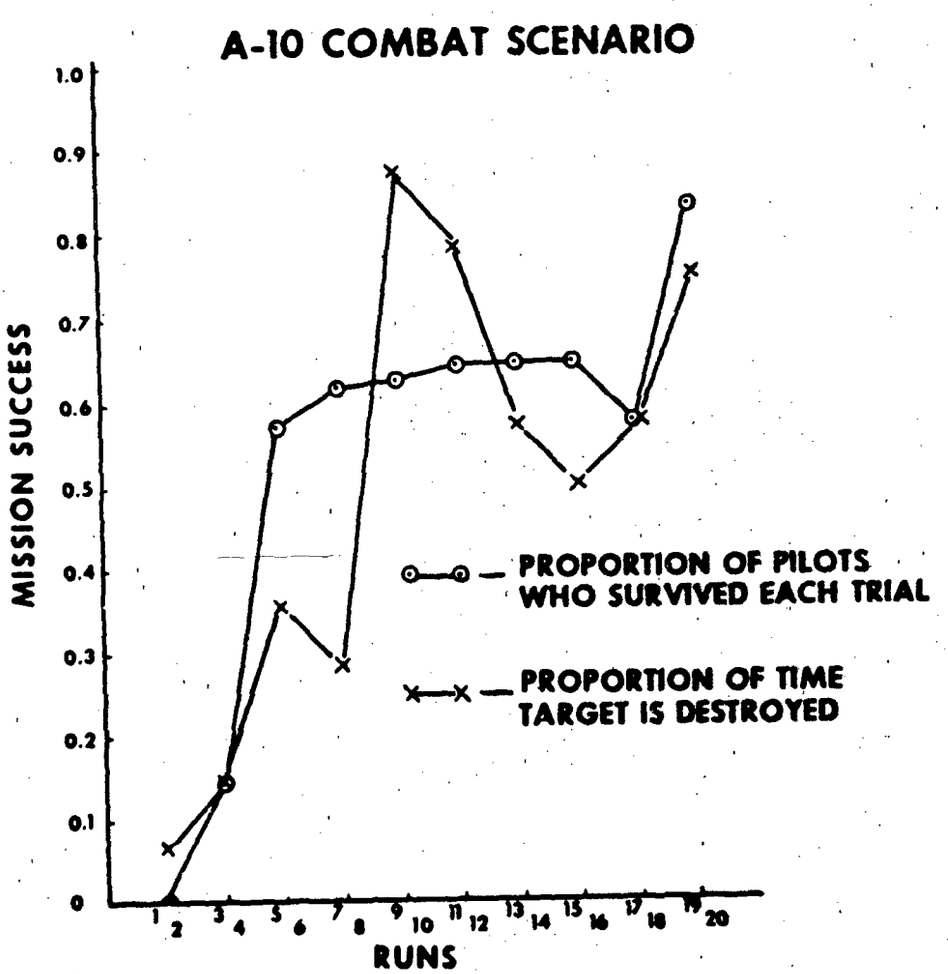


Figure 1

CONCLUSIONS

This being the first attempt to use the ASPT (or any other simulator) for the production of a full combat scenario, the preliminary results are quite encouraging. The visual system was programmed to almost full capacity to produce the battle environment in which the pilot was free to fly and develop his own tactics in destroying the target and surviving the threat environment. Thus, the pilot is operating in a completely open loop fashion, as he would largely operate in true combat and may be able to develop the flexibility and problem solving strategies he will be called upon to develop in actual combat. Learning curves indicate pilots are acquiring and integrating appropriate information to improve their tactical performance. The A-10 pilots' acceptance of this type simulation was remarkably positive. They pointed out that they had never had an opportunity, except for real combat, to deal with the combat environment in a holistic way as it was presented to them in the simulator. The demand of dealing with SAM and AAA weapons while conducting offensive maneuvers was particularly instructive. The "pop-up" maneuver, which is part of their standard repertoire, has to be clearly modified when in range of either the SAM or AAA site. They were in unanimous agreement that the simulation was very compelling and believable and with the use of wide screen visual presentation, G-seat and G-suit, platform motion was not necessary. In short, with the limited combined scenario presented, combat ready pilots exhibit clear learning of offensive and defensive tactics and favorable responses to such training.

RECOMMENDATIONS

It appears the simulator may be able to supply an important training methodology for tactics development. With development of greater computer capacity, visual and other systems may be capable of producing dynamic changing combat environments which would allow the combat ready pilot to maintain combat skills at a very high level. Flight simulation should fill the training void of those flying tasks which are impractical or impossible to practice in the real aircraft. It is therefore recommended that continued development of combat environments be made with ever increasing variability and flexibility so the total concept of simulator tactics development can be more thoroughly tested and exercised. It is also recommended as a follow-on, that comparisons be made between performance scores in exercises such as Red Flag with pilots trained in simulation and with pilots who have not had such exposure. Such a program could very well determine what sort of transfer of training is taking place from the simulator to the tactical exercise field.

REFERENCES:

1. Future Views: Aircrew Training 1980-2000. Memo by Maj Jack Thorpe, Life Sciences Directorate, Bolling Air Force Base, D.C. 20332

2. DeLeon, P. "The Peacetime Evaluation of the Pilot Factor in Air-to-Air Combat". Rand Corporation Report R2070-PA. Santa Monica, CA, 1977.

Developmental Strategy for Tactical Simulation of Ground Combat

David W. Bessemer
US Army Research Institute, Ft Knox Field Unit
Ft Knox, Kentucky 40121

Abstract

The conditions of the modern battlefield place heavy demands on leaders of platoons and weapons system crews for rapid assessment of tactical situations, effective decision-making, and efficient command, control, and communication (C^3). No method currently exists permitting small-unit leaders to practice critical cognitive skills in a realistic interactive environment representing the realtime interplay of important tactical factors. Technological developments in computer-generated imagery (CGI) and computer terminal networks have capabilities that might be developed to fill this need. Concepts for tactical simulation at four stages of complexity are described. Development of such prototype systems would permit examination of the tactical training potential of CGI without incurring the costs of full weapon system simulation.

Successful execution of the active defense in a European scenario implies a tactical perfection rarely attained in any previous Army. Consider the problem confronting a tank or mechanized infantry platoon leader and his subordinate vehicle commanders. They will be directly confronted by an enemy force of 10-50 armored vehicles, first detected at 2-4 km, and capable of closing on their position in 4 to 12 minutes. In this brief time, the platoon's firepower must be employed with maximum efficiency to reduce the enemy force, while minimizing exposure to enemy fire by use of terrain cover, and agile movement among firing positions. When the enemy force approaches to effective direct fire range, the platoon must rapidly move to a new battle position, avoiding being overrun, cutoff, or exposed to fire while in the open.

To conduct this battle the platoon leader must perform a host of tasks in rapid succession. He must report enemy contact, call for and adjust indirect fire, and continue to report the platoon situation, and execute directions from his company commander. He must decide when to open fire, maintain control of the distribution of fire, observe the effects of fire and enemy actions, hear reports of effects and actions, and decide whether to continue firing or to move to the next battle position. He must rapidly and correctly react to unexpected events, to adjust his plans to changed circumstances. At the same time, he must control his own vehicle's contribution to the fight.

From a psychological point of view, mastery of the information processing and decision-making demands implied by the execution of all of these duties in a brief time interval appears to be extremely difficult. It may be that a battle-hardened veteran can develop a well formed set of concepts, decision-rules, and automatic reactions that can handle the information-processing load. It is virtually certain that a green platoon leader will not have developed these cognitive skills to the level required.

PRECEDING PAGE BLANK-NOT FILMED

Tactical leadership at the small-unit level is clearly a key US Army training problem. Without extensive direct practice in a large variety of tactical situations under actual battle conditions, what training system can be devised to transform green platoon leaders and vehicle commanders into battle-hardened veterans before the first battle? In addressing this question, the capabilities and limitations of current training approaches will be briefly described, capabilities provided by generated CGI technology will be examined. Concepts will then be presented for a series of prototype systems that would permit the tactical training potential of CGI technology to be developed and evaluated.

Current Training Approaches. Tactical training systems now in use or under development support individual or team training for one or two crew members, large-unit command and staff exercises, or field exercises by small and/or large units. No current systems fully meet the needs of small unit tactical training.

The Conduct-of-Fire Trainers (COFTs) now under development for the XM1 and M60A1/3 tank systems will provide team training in gunnery and target engagement tactics for tank commanders and gunners. The COFTs provide a high-fidelity simulation of turret operations and fire control. A limited CGI system is used to present realistic threat arrays and tactics in the tank sights according to programmed scenarios, along with a variety of terrain and visibility conditions. Judicious use of the Control/Instructor station will provide limited capabilities to simulate some interactions with other crew members, actions of adjacent tanks, or communication with superiors. Potentially, COFTs could be linked together to simulate platoon action on common terrain, but this is unlikely to be pursued because of the costs associated with housing and maintaining multiple turret stations.

Engagement simulation using the new MILES laser equipment permits force-on-force field exercises with realtime casualty assessment and free command, control, and communication interaction. However, exercise costs and availability of maneuver areas seriously limit variety of terrain problems and opportunity for exercise repetition. The full array of threat weapons and tactics are not simulated, and the lack of scenario control permits unrepresentative situations to develop. Even the threat force and instrumentation planned for the National Training Center will not fully overcome these limitations.

Most of the tactical battle games and simulation systems now available in the US Army are designed to represent operations of the battalion or larger units. Only the Dunn-Kempf game provides tactical training at the platoon and company level. Dunn-Kempf uses a scaled terrain board and miniature models of vehicles along with game mechanics based on wargame hobbyist's rules. Weapon system capabilities and event sequences are reasonably approximated. Proper C^3 procedures are not required. Many aspects of the situation are not realistically represented, the most serious being the distortion of time scale. Up to eight hours of playing time are required to complete 10-15 minutes of combat. The players spend much of the time record-keeping and executing the game mechanics rather than practicing tactics. Tactical concepts and principles can be discovered and learned in Dunn-Kempf, but skill in realtime application of these concepts and principles must be gained in some other context.

Capabilities of CGI. CGI systems operate by computing a full-color picture based on a stored mathematical representation (data base) of a terrain area. Forms of surface vegetation, cultural objects, and moving or stationary vehicles and weapon systems can be reasonable approximated and included in the scene. The scene can be computed from any observation point, with correct perspective and occulting along the lines of sight, and proper magnification according to the sighting system.

Both the observation point and moving objects can move freely within the terrain area producing normal changes in perspective, occulting, and object size resulting from change in position. With sufficient computer resources, many vehicles can move over the terrain at the same time, and the same scene can be simultaneously viewed from many points of observation.

Numerous special effects can be integrated into the visual presentation. These include the effects of illumination intensity and direction, precipitation, smoke, dust, aerial perspective, weapon firing, and weapon effects. A large library of terrain and threat scenarios can be stored, providing nearly endless variety in training.

In essence, the CGI visual system provides a window on a simulated battlefield, with virtually complete flexibility and control of events unfolding on the battlefield in realtime. The simulated battlefield has a number of unique advantages for tactical training that cannot be duplicated in a field exercise environment. For example, a battle episode can be recorded, and then replayed repeatedly, to be viewed from any distance or direction selected at the time of the replay. No amount of videotaping in the field can provide this capability. A second kind of advantage is provided by the capabilities for special displays that can be overlaid on the battlefield scene for instructional purposes, such as a fire plan map showing fields of fire and intervisibilities from chosen firing positions.

Strategy for Research and Development. Until recently, a CGI simulated battlefield was under development as part of the Armor Full Crew Research Simulator (AFCRS). The AFCRS would have provided a total, high fidelity simulation of the XM1 tank, including driver and turret stations, motion platforms, CGI visual displays from all points of view, and much, much more.

Although the AFCRS project was terminated, some important lessons were learned. For example, the cost of developing and maintaining the crew station, motion hardware, and the associated computations were the biggest part of the package. Perhaps the most important lesson of the AFCRS is that hooking up full-crew simulators does not provide a viable concept for platoon tactical training, either in terms of cost or developmental risk.

For effective tactical training, a big investment in weapon system simulation does not seem to be really necessary. The cognitive skills required for tactical proficiency depend primarily on information input and feedback from the visual environment and various communication channels. Operating the weapon system machinery is not the core of the tactical problem, and might be minimized without much loss in training benefit.

Rather than building from the top down, as in the AFCRS, tactical simulation should be developed in building block fashion from the bottom up. If visual displays and communications are made available for leadership positions, we can learn gradually how to create realistic tactical problems, how to interface between stations, and what vehicle operations need to be

represented. As technical problems are overcome, additional stations can be added and the scope of the simulation increased. In this approach, various kinds of low-cost minicomputer and terminal networks using off-the-shelf components can be examined for suitability in meeting the display, control and computing requirements of tactical training. Following such an incremental strategy, both the development process and the resulting product should be cost-effective.

Tactical Simulation Systems. Concepts for tactical simulation systems are represented in Figure 1 at four increasing levels of complexity. The first level (Fig. 1a) is primarily intended to permit examination of the requirements for tactical data bases, and development of methods for presenting threat scenarios and special effects in realtime. The AVC station would provide an interactive visual display, and simple controls to initiate various actions. The control station would present visual displays from the AVC and enemy viewpoints, have manual driver and gunnery controls, and would provide control over the tactical scenario. Communication channels would permit simulation of interactions with the driver, gunner, or platoon leader from the control station. Movement and engagement decisions could be indicated directly on the AVC display, or through commands to the simulated crewmen. Ensuing actions and reactions would enfold at normal rates under automated or manual control. The CGI should be capable of presenting a company-sized threat (15 vehicles) in various 5 x 5 km terrain areas, from two viewpoints.

At the second level (Fig. 1b), two interacting leader positions are involved but the approach is generally the same. Simulated interactions with crewmen of both vehicles would be mostly automated, while communications from the control station simulate the subordinate AVCs and the company commander. About four viewpoints would be required, together with a battalion sized threat of about 35 vehicles in a 10 x 10 km area.

The complete platoon is represented in the third system (Fig. 1c). At this level of complexity about eight viewpoints are needed, along with a partial regimental threat of 60 vehicles in a 20 x 20 km area.

The platoon network and control station is used as a building block within the company network (Fig. 1d). If tactical simulation can be made to work effectively in this fourth stage, extensions of the building block approach are probably feasible to the battalion, brigade, and division levels. Interfacing the company network with existing battalion level simulations (such as CATTS) can also be attempted at this stage.

At any stage, the nature of the interactions between the control station and the leader positions can be examined to determine if direct interaction with gunner or driver is necessary in meeting the objectives of tactical training. If a requirement for interaction is established, the particular functions needed in gunner or driver stations can also be defined. Simple terminals, or low-fidelity partial crew stations could be quite sufficient for tactical training in each case, providing a substantial cost-reduction when compared to highly realistic crew stations.

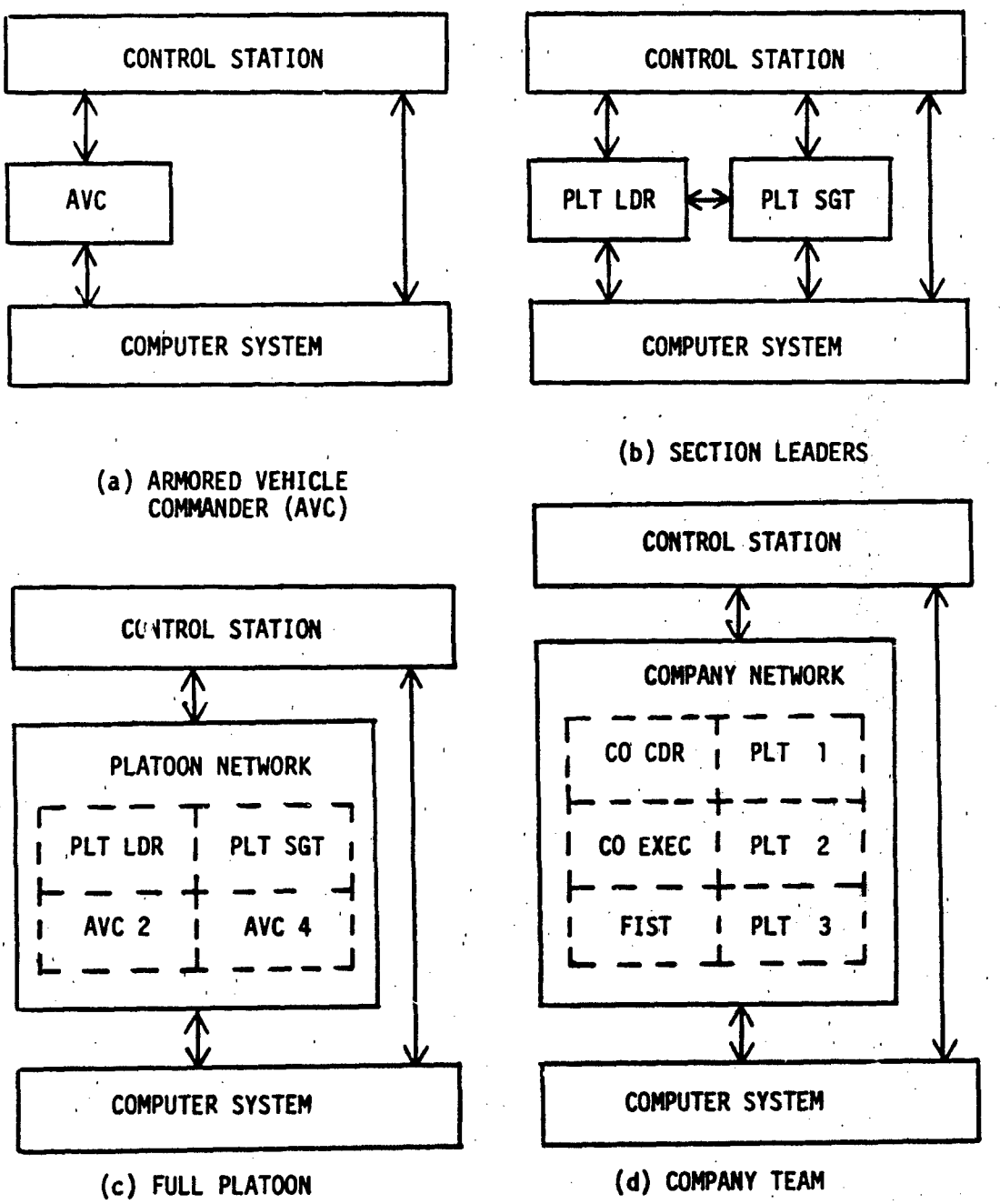


FIGURE 1. TACTICAL SIMULATION SYSTEMS

A FUNCTIONAL METRIC FOR DESCRIPTION OF VISUAL STIMULI

Russell A. Benel
University of Illinois at Urbana-Champaign

Abstract

A functional metric for stimulus adequacy based on the slope of the regression of accommodation (measured by a laser optometer) on stimulus presentation distance was tested. Observers (N=24) viewed four screens varying in contrast (3, 15, 75, and 95% contrast) at six optical distances (0-5 D). Other data were collected to determine the relationships among the point where the regression changed from lead to lag of accommodation and resting measures of accommodation. Slopes and intercepts of the regression differed among the screens and their correlation indicated that the regressions rotated about a point correlated with the dark focus. The relative independence of slopes from dark focus suggests the feasibility of further development of a functional metric that would yield direct information on stimulus adequacy.

Owens (1979) proposed quantifying stimulus effectiveness by optically varying the distance to the stimulus and simultaneously measuring accommodation. The slope of the regression of accommodation on stimulus presentation distance indexes the target's adequacy as an accommodative stimulus. Accordingly, slopes of 1.0 indicate veridical accommodation across object distances. Slopes of less than 1.0 reflect the presence of accommodative lead and lag even with high-contrast targets. Stimuli difficult to compare in spatial characteristics might be equated by slopes. The variety of parameters that characterize stimuli makes selection of descriptors difficult. Slopes would provide a common metric based on a response tendency.

Several issues relate to the use of slopes. Owens (1976) indicated interobserver (N=4) variability existed at each target condition, but the trends across observers were similar. To be most useful, slopes must not be idiopathic. If individual differences exist, the relation between response and some individual characteristic, e.g., the resting position for accommodation, is crucially important. Also, the crossover point referred to as the "fulcrum" for the accommodative stimulus-response relation (Johnson, 1976) has been described as the dark focus. Yet, reliable differences between the resting positions have been found (Benel & Benel, 1979). For example, the empty field condition more nearly resembles the viewing conditions when the stimulus has not been resolved, i.e., an apparently uniformly illuminated stimulus field.

Therefore, this experiment was designed to determine the following: (a) Can slopes of the regression of accommodation on presentation distance index the accommodative adequacy of stimuli? (b) Are slopes and resting measures independent, and, if not, are there lawful relationships between them such that slopes can still be useful? (c) What is the relationship between resting position measures? (d) Which resting position measure (if they differ) is most representative of the accommodative fulcrum?

Method

Observers. The 24 nominally emmetropic observers (aged 18-30) had near

PRECEDING PAGE BLANK-NOT FILMED

and far visual acuities of at least 20/25 measured by an Orthorater.

Stimuli. The prototype screen stimulus consisted of crossed rectilinear strands subtending 7.2 min visual angle (VA) separated by 16.8 min VA. Four screen stimuli were produced by placing high-contrast photographic positives (black lines on a clear background) at various distances from matte diffusing surfaces. The non-blurred stimulus placed the transparency on the side of the matte material toward the observer. The percent contrast associated with each stimulus (3, 15, 75 and 95%) was computed according to the following fomula: Contrast (%) = 100(LB-LT)/LB. The background (LB) was the light area and the target (LT) was the dark area.

Apparatus. Stimuli were presented in a two channel Maxwellian view optical system (see Fig. 1). A Sawyer projector (model 500XM) with CWD projection bulb (120 V, 300 W) served as the light source. Lenses L1 and L2 formed a bright field on the opal glass screen (OG1) masked by a field stop (FS1) to form a 14 mm circular image at the observer's entrance pupil. Stimuli were positioned in the collimated portion of Channel 2 between lenses L5 and L6 which are 180 mm focal length yielding a maximum dioptric power of 5.56 D. The 12 deg circular stimulus field was limited by a field stop (FS2) placed at -5.56 D (beyond optical infinity). A severely out of focus image does not act as an accommodative stimulus (Smithline, 1974).

Accommodation was measured with a laser optometer (Hennessy and Leibowitz, 1972). The beam of a 2.0 mW He-Ne laser (Metrologic Model MC-650) was diverged (L7), collimated (L8), and then reflected by a mirror (M1) from the surface of a slowly rotating drum (RD). The resulting speckle pattern (10 deg VA) was superposed on the observer's field of view (BS1) and the intensity was adjusted (FH2) until only the brightest speckles remained visible. Exposure duration (0.5 sec out of 5.0 sec) was controlled by a rotating beam chopper. Speckle movement indicated the observer's refractive state. If overaccommodated for the test pattern, the speckles "flow" with the drum's rotation; if underaccommodated, they flow opposite the rotation. When the "plane of stationarity" (Charman, 1974) is conjugate with the retina, the speckles appear stationary or merely swirling. Bracketing movements are made with the drum to locate this plane. The insertion of a positive lens in the light path of the laser pattern one focal length from the observer's entrance pupil allows the plane of stationarity to be varied from nearly the dioptric power of the lens to beyond optical infinity with no changes in the brightness or size of the test pattern (Ogle, 1971, p. 226). An additional correction (.32 D) is added for the monochromatic light of the He-Ne laser (633 nm).

A dental impression bite-board adjustable in X, Y, and Z planes held the observer's head in proper position. Screens of black construction board prevented observation of the apparatus during data collection. The room lights were off during data collection.

Procedure. The observer was seated and aligned with the apparatus so that the stimulus field appeared centered and maximally bright. After instruction in the use of the laser optometer and several familiarization trials, the observer's resting measures were taken with and without the Maxwellian view illuminated. Two consecutive measures of each resting state were collected. Next, the screens were presented in counterbalanced order at optical distances of 0 through 5 D in 1 D increments. At each stimulus-distance combination two successive accommodation measures were

taken. Observers were not informed of the stimulus distances and were reminded to observe the presented stimulus carefully.

Results

The effects of screen type and distance upon accommodation were analyzed by a 4-way ANOVA (see Table 1). Because the variance associated with the replication factor was spuriously small, F-ratios using this term as the denominator were not constructed. The main effects for both screens and distances and their interaction were reliable. Accommodation became less accurate when stimulus contrast decreased. Furthermore, the inaccuracy is greatest at the extremes of the range tested. Accommodation changed very little with stimulus distance for screen 4.

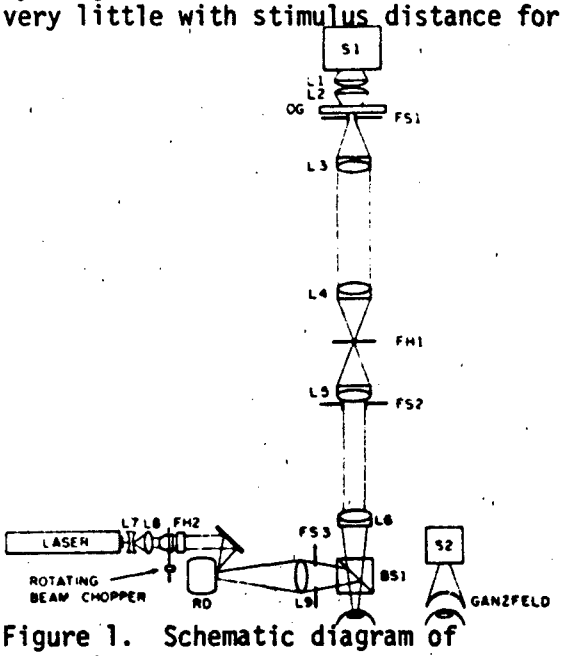


Figure 1. Schematic diagram of apparatus.

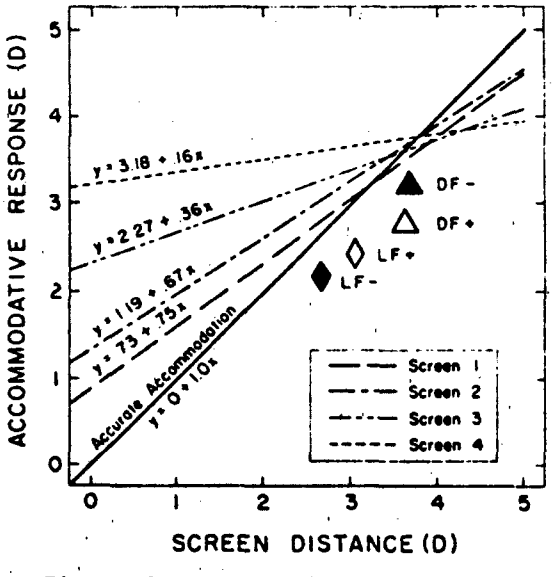


Figure 2. Regressions for the four screen stimuli.

Table 1

Analysis of Variance for Accommodation to the Screens

Source	SS	DF	MS	F	ω^2
Screens (A)	114.33	3	38.110	7.04**	.09
A X C	352.65	69	5.111		
Distances (B)	840.13	5	168.026	126.72***	.54
B X C	152.49	115	1.326		
A X B	243.07	15	16.205	19.94***	.22
A X B X C	280.35	345	.813		
Observers (C)	496.32	23	21.582		
Replications (D)	21.74	576	.038		

** $p < .01$ *** $p < .0001$

The procedure to assess the strength of the association between the treatments and the dependent variable, followed that outlined by Vaughn and Corballis (1969). The ratio is equivalent to, but slightly underestimates, what is known as omega squared (ω^2). The proportion of variance accounted for by the treatments totalled 85%.

Least-squares linear regressions were computed for the mean accommo-

dative response on stimulus presentation distance. Two 2-way ANOVAs (see Tables 2 and 3) were performed on the slopes and intercepts. Both differed reliably across screens. The proportion of variance due to screens in the two analyses was 49 and 52% respectively. The mean intercepts and slopes differed reliably (Tukey test, $p < .05$) with the exception of screens 1 and 2 (and those results were in the expected direction). The negative Pearson product-moment correlation between slopes and intercepts supported the contention that the regression line rotates about a fulcrum ($r(24) = -.75$, $p < .01$) (see Fig. 2).

Tables 2 and 3

Analysis of Variance for the Intercept of the Screens

Source	SS	DF	MS	F	ω^2
Screens (A)	87.253	3	29.0861	31.52***	.49
A X B	63.6664	69	.9227		
Observers (B)	63.704	23	2.7699		
Analysis of Variance for the Slope of the Screen					
Screens (A)	5.4496	3	1.8165	34.31***	.52
A X B	3.6531	69	.0529		
Observers (B)	2.404	23	.0144		

*** $p < .001$

The relationship between the fulcrum and resting positions was examined by Pearson product-moment correlations with the median crossover point (where the regression lines crossed the line of accurate accommodation). The correlation ($r(24) = .91$, $p < .05$) with the dark focus was reliable, but that with the light focus was not.

Discussion

Accommodative responses to the high-contrast screen were reasonably accurate. Consistent with the findings of Owens (1979) and others (e.g., Charman & Tucker, 1977, 1978) some lead of accommodation was seen at the most distant position, but the average observer exhibited relatively little lag until the near stimulus position (5 D). As contrast decreased, accuracy decreased when the stimulus position departed in either direction from the observer's resting accommodation distance. Comparison of the derived slope for accommodative functioning and the objective change in stimulus characteristics (characterized as % contrast) indicates the plausibility of this technique for the functional description of stimulus adequacy. Although, a wider range of objective characteristics need to be investigated prior to total acceptance, this is, nevertheless, an important initial step for determining the generality and applicability of this metric.

Individual differences did not appear to play a major role in determining the direction of the accommodative response within this experiment. The proportion of variance accounted for by treatment effects left only 15% to be attributed to observers and other sources. Likewise, slope and intercepts appeared to be relatively independent of individual differences in the dark focus for the screens of high contrast. The higher correlations between the intercepts and the dark focus found when visual resolution of the stimulus becomes more difficult reflected the passive return toward the resting position. Apparently the absolute level of the accommodative response depended more on the dark focus than on the actual stimulus distance as contrast decreased.

The dark focus appeared to relate most closely to the accommodative

fulcrum. Although this was consistent with previous research, the lower correlation between the light focus and intercept for the low-contrast stimulus was not expected. Apparently the unresolved, low-contrast stimulus is not identical to the empty stimulus field. The absolute difference between the dark focus and light focus in this experiment was somewhat larger than that reported by Leibowitz and Owens (1978). Perhaps the particular viewing system used affected the results and a pure Ganzfeld might yield higher correlations, but the dark focus appears to predict the crossover accurately.

Acknowledgement

This research was supported under AFOSR contract USAF F49620-77-C-0117 with the U. of IL at Urbana-Champaign. Dr. A. R. Fregly was the scientific monitor and Dr. S. N. Roscoe the principal investigator. This research was part of a dissertation submitted in partial fulfillment of the requirements for the Ph.D. The author is now a National Research Council Postdoctoral Associate at USAFSAM/VNE, Brooks AFB TX 78235.

References

Benel, R. A. and Benel, D. C. R. Accommodation in untextured stimulus fields. Champaign, IL: Department of Psychology, U. of Illinois at Urbana-Champaign, Technical Report ENG PSY-79-1/AFOSR 79-1, 1979.

Charman, W. N. On the plane of stationarity in laser refraction. American Journal of Optometry and Physiological Optics, 1974, 51, 832-838.

Charman, W. N. and Tucker, J. Dependence of accommodation response on the spatial frequency spectrum of the observed object. Vision Research, 1977, 17, 129-139.

Charman, W. N. and Tucker, J. Accommodation as a function of object form. American Journal of Optometry and Physiological Optics, 1978, 55, 84-92.

Hennessy, R. T. and Leibowitz, H. W. Laser optometer incorporating the Badal principle. Behavioral Research Methods and Instrumentation, 1972, 4, 237-239.

Johnson, C. A. Effects of luminance and stimulus distance on accommodation and visual resolution. Journal of the Optical Society of America, 1976, 66, 138-142.

Leibowitz, H. W. and Owens, D. A. New evidence for the intermediate position of relaxed accommodation. Documenta Ophthalmogica, 1978, 46, 133-147.

Ogle, K. N. Optics: An Introduction for Ophthamologists. Springfield, IL: Thomas, 1971.

Owens, D. A. Factors influencing steady-state accommodation. Unpublished Ph.D. Dissertation. Pennsylvania State University, November 1976, University Microfilms, Ann Arbor Michigan.

Owens, D. A. The Mandelbaum effect: Evidence for an accommodative bias toward intermediate viewing distances. Journal of the Optical Society of America, 1979, 69, 646-651.

Smithline, L. M. Accommodation responses to blur. Journal of the Optical Society of America, 1974, 64, 1512-1516.

Vaughn, G. M. and Corballis, M. C. Beyond tests of significance: Estimating strength of effects in selected ANOVA designs. Psychological Bulletin, 1969, 72, 204-213.

Cognitive Maps and the Subjective Magnitude of Large-Scale Distances

Edward J. Rinalducci
School of Psychology
Georgia Institute of Technology
Atlanta, GA 30332

and

Douglas A. Hardwick
Department of Psychology
Illinois State University
Normal, ILL 61761

Abstract

The question asked by this study is that if people have different organizations of cognitive maps as suggested by Moore, do they also perceive the distances between elements of the environment such as buildings or landmarks differently. Two levels of cognitive map organization were derived from the subjects' sketch maps of a university campus. Subjects also estimated distances between selected buildings using the free-modulus variation of the method of magnitude estimation. No differences were obtained between the different levels of cognitive map organization. However, significant differences were found when distance estimates were made from the subjects' actual location vs. an imagined location elsewhere on campus. Estimates made from an imagined locus were closer to the real distances, suggesting that subjects were forced to use a higher level topographical representation.

A cognitive map is a theoretical construct which is inferred from human spatial behaviors. A cognitive map is assumed to be the product of some unspecified process by which an individual encodes and stores spatial information from the environment (Blaut, McCleary, & Blaut, 1970). It may be useful to consider a cognitive map as a large-scale mental image, which is specific to some portion of the environment. However, a cognitive map can also be conceptualized simply as an individual's organized information about the environment's spatial layout, i.e., information about distance, directionality, and alternate routes among specified landmarks. In this conceptualization, the mental image becomes a "pictorialization" (Downs & Stea, 1973). The power of the cognitive map construct derives from its assumed role in the determination of human spatial behaviors and decisions. In this context, an analogy is often drawn between a cognitive map and the common road map (Hardwick, McIntyre, & Pick, 1976). A cognitive map and a road map are both viewed as functioning to facilitate an individual's orientation and movement within the environment. To the extent that a cognitive map is accurate, the behaviors mediated by that map will also tend to be accurate and efficient. If a cognitive map is inaccurate, the individual's behaviors will be inefficient and disoriented (Harrison & Howard, 1972).

The approach to the study of cognitive maps has typically been through the use of map drawing tasks (Lynch, 1960) or through distance estimation (Briggs, 1973; Cadwalleder, 1976). The present study is a combination of both approaches. In particular, this study seeks to examine the relationships between cognitive map organization and the subjective magnitude of large-scale distances.

One of the most thoughtful analyses of how cognitive maps are organized has been provided by Moore (1973, 1974, 1976). Integrating both child and adult research, Moore believes that cognitive maps pass through developmentally ordered stages of organization. Moore has identified three levels of organization. Each of these levels is discussed briefly below.

At Level I, a cognitive map is an action-based representation tied to one or two salient experiences within the environment. Such a map does not organize an environmental unit as a whole. Only topological relations (e.g., proximity, order) are preserved in cognitive maps at this level. The sketch maps which result are egocentric and concrete, e.g., they contain picture-like drawings of landmarks encountered along a favorite pathway. They are essentially route maps.

At Level II, the cognitive map of a given environmental unit is organized in terms of "clusters" of elements. The elements within such clusters are systematically related to one another, e.g.,relative distances may be preserved within a cluster. The clusters themselves, however, are not systematically related to one another. The sketch maps which result are characterized by a higher degree of integration and organization within clusters than across clusters.

At Level III, a cognitive map is a single, coordinated representation of the entire environmental unit. Elements of the representation are related to one another by means of a comprehensive system of reference. It is this type of organization which permits systematic cognitive operations on the entire representation.

According to Moore, the progression from Level I to Level III cognitive maps can occur microgenetically as well as ontogenetically. This position would account for the presence of route-based organizations in sketch maps of both children and adults. In addition, it is consistent with a shift toward spatially dominant sketch maps as a function of length of residence (Appleyard, 1970).

The question asked by this study is that if people have different organizations of cognitive maps (given a certain amount of experience with a particular environment) as suggested by Moore, do they perceive the distances between elements of the environment such as buildings or landmarks differently?

Method

Subjects. The subjects were 34 undergraduate students enrolled in introductory psychology courses at the University of

Virginia. All subjects had attended this institution for at least a year prior to being in the study. The subject's participation fulfilled a course requirement.

Procedure. Three variables were examined in this study. They included order of experimental session (drawing of sketch map first and making distance estimates second or vice versa), map level (Moore's Levels II and III), and locus (the actual location of the subject vs. an imagined location elsewhere on the grounds).

One experimental session was used to collect sketch maps of the grounds of the University of Virginia, and the other to obtain estimates of distances between buildings located on the grounds. In order to control for possible order effects deriving from drawing sketch maps first and then estimating real-world distances or vice versa, two groups of subjects were run. The first group (19 subjects) drew sketch maps first and estimated distances second. The second group (15 subjects) made distance estimates in the first session and sketch maps in the second.

Cognitive map organization was determined for all 34 subjects using the criteria developed by Moore (1973, 1974, 1976). Three judges, including the two authors and a research assistant, judged the sketch maps independently for level of organization. Only Level II and III maps were obtained. The assigned level of organization was determined by majority agreement.

Subjects were instructed in the free-modulus variation of the method of magnitude estimation (Engen, 1972). In order to familiarize the subjects with the technique, they were requested initially to estimate the length of lines. The subjects were then presented with the names of 8 buildings (in a random order) located on the grounds of the University of Virginia. Subjects were requested to estimate distances to the various buildings using their own actual location (i.e., the building in which they were making the distance judgments). The distances from the subjects' actual location varied from about 183 m for the nearest building to 899 m for the farthest building. The buildings were distributed, more or less, in the same general direction from the actual location to the farthest of the 8 buildings. The subjects were then asked to imagine they were located in the farthest building and to estimate the distances back to their actual location. Again the same 8 buildings were involved in the distance estimations. The subjects were acquainted with the general location and purpose of all the buildings. In all cases, whether it was estimating length of lines or distances between buildings, two trials were given for each stimulus. All stimuli were projected onto a screen.

Results

Each of the subjects was familiar with the grounds of the University of Virginia and had been in residence for at least

a year. Therefore, it was not surprising to find that none of the sketch maps were assigned Level I. A total of 16 subjects were assigned a sketch map organization of Level II and 18 were assigned Level III. Of the 16 with Level II maps, 9 were in the group which drew sketch maps first and 7 were in the group which estimated distances first. Of the 18 assigned Level III, 10 drew sketch maps first and 8 estimated distances first.

Exponents were derived for each subject from the regression line of the log value of the actual line length vs. the log value of estimated line length. Each exponent was treated as an individual data point. These data were then analyzed using a 2 x 2 ANOVA (Level x Order) experimental design. No significant differences were obtained. The overall value of the exponent was found to be 1.046, which is quite close to the value of unity reported by Stevens (1961). An exponent of 1 indicates that an increase in actual line length led to a proportional increase in estimated length.

Exponents were also determined for each subject from a regression line of actual vs. estimated real-world distances made from the subjects' actual location and from an imagined location. These data were analyzed using a 2 x 2 x 2 ANOVA (Level x Order x Locus) experimental design. The only significant difference obtained was for the variable of locus (actual vs. imagined location) where a F equal to 6.01 (degrees of freedom equal to 1 and 64) was obtained with p less than 0.05. The overall exponent for actual location estimates was 0.714 while that for the imagined location was 0.826. This result indicates that the value of the exponent for the imagined location approaches unity more closely than those estimates made using the actual locus.

Discussion

The results of this study indicate that the level of cognitive map organization demonstrated by a subject's sketch map does not seem to be related to the subject's ability to estimate real-world distances. This result does not support Moore's hypothesis concerning the organization of mental maps. Perhaps a more stringent criterion in the assignment of level of organization (i.e., unanimous agreement) would have provided different results. Certainly future work along the same lines should incorporate such a criterion.

A larger exponent was obtained when subjects estimated real-world distances using an imagined locus as compared to their actual location. By having subjects use an imagined locus perhaps we are forcing them to use a higher level topographical representation, and thus approach an exponent of 1 more closely than if the actual location was employed. When the actual location is used, the subject may be using a lower level map and has to estimate across clusters. By estimating across clusters of Moore's Level II cognitive map, the subject

may be leaving out some of the actual distance between them. The underestimation of distance should result in a lower exponent. These results suggest that if we want to improve an individual's ability to estimate real-world distances we must train him or her to use an imagined rather than an actual locus.

References

Appleyard, D. Styles and methods of structuring a city. Environment and Behavior, 1970, 2, 100-108.

Blaut, J. M., McCleary, G. F., & Blaut, A. S. Environmental mapping in young children. Environment and Behavior, 1970, 2, 335-349.

Briggs, R. Urban cognitive distance. In R. M. Down & D. Stea (Eds.), Image and environment. Chicago: Aldine, 1973, 361-388.

Cadwallader, M. T. Cognitive distance in intraurban space. In G. T. Moore & R. G. Golledge (Eds.), Environmental knowing. Stroudsburg: Downden, Hutchinson & Ross, 1976.

Down, R. M., & Stea, D. (Eds.). Image and environment. Chicago: Aldine, 1973.

Engen, T. Psychophysics II. scaling methods. In J. W. Kling & L. A. Riggs (Eds.), Woodworth & Schlosberg's experimental psychology, vol. I: sensation and perception. New York: Holt, Rinehart, and Winston, 1972, 47-86, 3rd ed.

Hardwick, D. A., McIntyre, C. W., & Pick, H. L. The content and manipulation of cognitive maps in children and adults. Monographs of the Society for Research in Child Development, 1976, 41, (3).

Harrison, J. D., & Howard, W. The role of meaning in the urban image. Environment and Behavior, 1972, 4, 389-411.

Lynch, K. The image of the city. Cambridge, Mass.: Harvard University Press and M.I.T. Press, 1960.

Moore, G. T. Conceptual issues in the study of environmental cognition: an introduction. In W. J. Mitchell (Ed.), Environmental design: research and practice, vol. 2. Los Angeles: School of Architecture and Urban Planning, University of California, 1972, pp 30-1-1 to 30-1-3.

Moore, G. T. Developmental differences in environmental cognition. In W. F. E. Preiser (Ed.), Environmental design research, vol. 2. Stroudsburg: Dowden, Hutchinson & Ross, 1973, 232-239.

Moore, G. T. The development of environmental knowing: an overview of an interactional constructivist theory and some data on within-individual developmental variations. In D. Canter & T. Lee (Eds.), Psychology of the built environment. New York: John Wiley, 1974, 184-194.

Stevens, S. S. The psychophysics of sensory function. In W. A. Rosenblith (Ed.), Sensory communication. Cambridge, Mass.: The M.I.T. Press, 1961, 1-33.

Eye Accommodation, Personality, and Autonomic Balance

Valerie J. Gawron
Behavioral Engineering Laboratory
New Mexico State University
Las Cruces, New Mexico 88003

Abstract

The relationships among autonomic balance (as measured by a battery of four physiological tests modified from Wenger and Ellington, 1943, and by a technique introduced by Porges, 1976), refractive error (measured by dark focus, near and far points using a polarized vernier optometer), and introversion-extraversion (Eysenck Personality Inventory introversion-extraversion scale score) were investigated. It was expected that Pearson product-moment correlation coefficients among these measures would constitute a critical test of the validity of existing theoretical and empirical conclusions. The subjects of this study were 152 military recruits. The results indicate that the relationships among the variables were not as predicted by theory and previous research.

The autonomic nervous system is made up of two subsystems: the parasympathetic (PNS) and the sympathetic (SNS). The balance between these systems regulates bodily functioning during routine (PNS-dominant) and crisis (SNS-dominant) situations. It also controls visual accommodation for near (PNS-dominant) and far (SNS-dominant) focus (Cogan, 1937; Mohney, Morgan, Olmsted, & Wagman, 1942; Olmsted, 1944).

The balance between these physiological systems has been linked to individual differences in personality characteristics, especially introversion (PNS-dominance) and extraversion (SNS-dominance) (Eysenck, 1953). Since the balance mediates accommodation, the similar personality differences between near- and far-sighted individuals (Beedle & Young, 1976; Mull, 1948; Randle cited in Roscoe & Benel, 1978) may be related to the more general parasympathetic-sympathetic balance rather than being related solely to the visual capability difference. The purpose of this research was to examine these relationships.

Method

Autonomic Balance

Wenger's $\bar{A}$. A modified version of the Wenger and Ellington (1943) $\bar{A}$-battery was used to produce one index of autonomic balance, $\bar{A}$, that varies from low scores (SNS-) to high scores (PNS-dominance). $\bar{A}$-scores form a normal distribution across subjects, remain relatively constant over time for any one subject and are both reliable and valid (Wenger, Engel, & Clemens, 1957).

Three of the seven physiological measures that made up the original battery were excluded since their validity was not demonstrated in the psychobiological literature. The following were included, however:

Heart period. Cardiac activity was measured with Beckman standard-size Ag/AgCl electrodes filled with Beckman electrode paste and attached to the volar surface the right forearm and calf of the left leg with electrode

PRECEDING PAGE BLANK-NOT FILMED

collars. An electrode on the right calf served as ground. EKG was monitored with a Beckman A-C 9806A coupler set in a Beckman RB Dynagraph and a 2-minute recording was made using an Ampex SP 700 FM tape recorder. Heart period was scored in milliseconds as the mean interbeat interval over 110 seconds.

Skin conductance. Beckman standard Ag/AgCl electrodes were attached to the pads of the middle fingers of the right hand with electrode collars and surgical K-Y jelly. SC was recorded (1 mm = 1 micromho) relative to a constant 0.5 volt DC reference imposed by a Beckman 9842 Galvanic Skin Response Coupler. Two measurements, each of one-minute duration, were made.

Respiration period. A rubber bellows was placed on the mid-thorax. Pressure changes occurred in the bellows due to breathing movements, were transduced to voltage by a Grass Model PT 5-A volumetric pressure transducer, and monitored by a second Beckman 9806A A-C coupler. Respiratory activity was recorded in the same manner as heart period. The reciprocal of the dominant respiratory frequency was the same for this test.

Pulse pressure. Diastolic pressures were measured by auscultation at the brachial artery of a subject with cuff on the upper left arm. Six readings were taken over a 20-minute interval. The score was the mean diastolic reading, expressed in mm Hg, for the two lowest systolic pressures that duplicated each other.

Porges' C_W. This estimate of autonomic balance is described by Porges (1976) and based on heart and respiration activity. Cross spectral analysis of simultaneous recordings of these activities produces a coherence statistic (a normalized function with values ranging between zero and one) that may be an estimate of central autonomic functioning. C_W has been shown to be both reliable and valid (Porges, 1976; Porges, Coles, Drasgow, & Bohrer, 1978).

Introversion-Extraversion

Eysenck Personality Inventory. This inventory contains a reliable and valid introversion-extraversion (I/E) scale (see Gawron, 1979). High scores on this scale indicate a tendency towards extraversion.

Visual Accommodation

The following were measured using a polarized vernier optometer as described by Simonelli (1979b): a) dark focus or resting accommodation, the dioptric power of the eye in the absence of an external stimulus, b) near point, the limit for positive accommodation and thus the theoretical maximum of PNS influence on eye accommodation (near point was also measured using an RAF Near Point Rule), c) far point, the limit for negative accommodation and the point of maximum SNS influence, and d) dark focus range.

Subjects

One hundred fifty-two military trainees between the ages of 18 and 30 were the subjects of this experiment. All participated voluntarily.

Procedure

A subject completed the EPI and then viewed a polarized vernier optometer to determine dark-focus, near- and far-point accommodations. Electrodes were then attached and EKG, skin conductance, and respiration recorded as previously described.

Results

The physiological recordings of 27 subjects were eliminated from the data base due to the poor quality of these recordings. For the remaining 135 subjects, $\overline{A}$ was determined in accordance with Wenger and Ellington (1943) as the sum of the products of the absolute value of the factor scores (from a two-factor, principal components factor analysis) for heart period, skin conductance, respiration period, and pulse pressure times each subject's T-standardized score on the respective variable.

C_w was calculated by the following procedure: heart period was sampled from the data tape every 250 msec and was computed as the sum of each heart period that partially or wholly occupied the 250-msec interval. These computations were multiplied by the proportion of the interval in which the heart period was represented. Respiration amplitude was also sampled from the data tape every 250 msec. The time series for heart period and respiration amplitude were then prestationed by removing linear trends and the mean (Porges, Bohrer, Keren, Cheung, Drasgow, & McCabe, 1979). The data were cross-spectrally analyzed and a weighted coherence coefficient calculated.

A Pearson product-moment correlation matrix based on each individual's scores on $\overline{A}$ and C_w, the I/E scale of the EPI, as well as his or her mean dark focus over three trials, range of dark focus measures, and near and far points formed the data for the matrix presented in Table 1. Reliable correlations existed between $\overline{A}$ and several other measures. $\overline{A}$ and C_w were positively correlated (r = +.20) yielding consensual support. $\overline{A}$ was negatively correlated with both dark focus (r = -.21) and far point (r = -.18) indicating that PNS-dominant individuals have farther dark focuses and far points (SNS characteristics). However, C_w was positively correlated with near point (r = +.18) implying that PNS-dominant individuals have nearer near points. This last finding supports the continuity hypothesis (Porges, 1976) while those of $\overline{A}$ with dark focus and far point are contradictory to it. Also, extraverted subjects had smaller ranges of dark focus (r = -.16) than more introverted subjects.

Table 1

Correlation Matrix of Baseline Physiological and Personality Measures

	$\overline{A}$	C_w	I/E	DF	DFR	NP
$\overline{A}$ (Wenger's index)						
C_w (Porges's index)	.20*					
I/E (Eysenck)	.10	.05				
Mean Dark Focus (DF)	-.21*	-.15	.03			
Dark-Focus Range (DFR)	-.01	-.06	-.16*	.07		
Near Point (NP)	-.04	-.18*	.00	.26**	-.10	
Far Point (FP)	-.18*	-.12	.06	.94**	-.03	.29**

*p < .05; **p < .01

Discussion

Eysenck (1953) suggests that PNS-dominant individuals are introverted while SNS-dominant individuals are extraverted. This hypothesis would have been supported if there were reliable negative correlations between $\bar{A}$ or C_w and I/E scores. No such correlations were found. Nor were skin conductance and I/E scores reliably correlated. This finding coincides with those of Burdick (1966), Purohit (1966), Revelle (1974), and Small (1976). However, I/E was related to dark focus range; introverted subjects tended to have larger ranges of dark focus.

Randle (cited in Roscoe & Benel, 1977) proposed that PNS-dominant individuals tend to be myopic, SNS-dominant individuals, hyperopic. This hypothesis would have been supported if and only if $\bar{A}$ were positively correlated with near point, dark focus, and far point and negatively correlated with dark focus range. $\bar{A}$ was reliably correlated to dark focus and far point but in the direction opposite to that proposed by Randle. This suggests that it was the SNS-dominant individuals who were more myopic. However, the relationship between C_w and range of dark focus was both reliable and in the direction predicted by Randle. C_w was also reliably correlated with near point indicating that, as Randle suggested from personal observation, PNS-dominant individuals have nearer near points.

Most researchers have found myopia to be related to introversion. This finding would have been replicated if there were strong, negative correlations between scores on the I/E scale of the EPI and dark focus and near and far points. Although the ranges of visual accommodation and I/E scores were broad, none of these correlations was reliable. However, researchers who have found myopes to be introverted used tests other than the EPI to measure introversion. Mull (1948) used the Bernreuter Personality Inventory while Beedle and Young (1976) used the Omnibus Personality Inventory. Since the EPI scale measures did not yield results comparable to others found in the literature, it may be a poor measure of introversion. Conversely, it may measure something which the other tests do not. As often occurs in psychology, constructs with the same name may not be equivalent.

Also, Mull (1948) and Beedle and Young (1976) used college students as subjects. These individuals tend to be more myopic than the general population. Indeed over half of Mull's subjects were myopes. In the present study, a very different population was used, military recruits (see Simonelli, 1979a). The difference in results may reflect the different data bases.

References

Beedle, S. L., & Young, F. A. Values, personality, physical characteristics and refractive error. American Journal of Optometry and Physiological Optics, 1976, 53, 735-739.

Burdick, J. A. Autonomic lability and neuroticism. Journal of Psychosomatic Research, 1966, 9, 339-342.

Cogan, D. G. Accommodation and the autonomic nervous system. Archives of Ophthamology, 1937, 18, 739-766.

Eysenck, H. J. The structure of human personality. London: Metheun, 1953.

Gawron, V. J. Eye accommodation, personality, and autonomic balance (BEL-79-2/AFOSR-79-6). Las Cruces, NM: New Mexico State University, Behavioral Engineering Laboratory, November 1979.

Mohney, J. B., Morgan, M. W., Olmsted, J. M. D., & Wagman, I. H. The pathway of sympathetic nerves to the ciliary muscles in the eye. American Journal of Physiology, 1942, 135, 759-762.

Mull, H. K. Myopia and introversion. American Journal of Psychology, 1948, 61, 575-576.

Olmsted, J. M. D. The role of the autonomic nervous system in accommodation for near and far vision. Journal of Nervous and Mental Disease, 1944, 99, 794-798.

Porges, S. W. Peripheral and neurochemical parallels of psychopathy: A psychophysiological model relating autonomic imbalance to hyperactivity, psychopathy, and autism. Advances in Child Development and Behavior, 1976, 11, 35-65.

Porges, S. W., Bohrer, R. E., Keren, G., Cheung, M. M., Drasgow, F., & McCabe, P. A new time-series statistic for detecting rhythmic co-occurrence in the frequency: The weighted coherence and its application to psychophysiological research. Unpublished paper, University of Illinois at Urbana-Champaign, 1979.

Porges, S. W., Coles, M. G. H., Drasgow, F., & Bohrer, R. Individual differences in cardiac respiratory coupling and heart rate response during two attention demanding tasks. Paper presented at the Annual Meeting of the Society for Psychological Research. Madison, Wisconsin, September 1978.

Purohit, A. P. Personality variables, sex difference, GSR responsiveness, and GSR conditioning. Journal of Experimental Research in Personality, 1966, 1, 166-173.

Revelle, W. R. Introversion/extraversion, skin conductance and performance under stress. Dissertation Abstracts International, 1974, 35, 487-488.

Roscoe, S. N., & Benel, R. A. Is the eye smart or the brain forgiving? (Eng Psy-78-1/AFOSR-78-1). Urbana-Champaign, IL: University of Illinois, Department of Psychology, April 1978.

Simonelli, N. M. The dark focus of visual accommodation: Its existence, its measurement, its effects (BEL-79-3/AFOSR-79-7). Las Cruces, NM: New Mexico State University, Behavioral Engineering Laboratory, November 1979a.

Simonelli, N. M. Polarized vernier optometer (BEL-79-4/AFOSR-79-8). Las Cruces, NM: New Mexico State University, Behavioral Engineering Laboratory, November 1979b.

Small, C. S. Autonomic response specificity and extraversion. Journal of Clinical Psychology, 1976, 32, 537-538.

Wenger, M. A., & Ellington, M. The measurement of autonomic balance in children: Method and normative data. Psychosomatic Medicine, 1943, 5, 241-253.

Wenger, M. A., Engel, B. T., & Clemens, T. L. Studies of autonomic response patterns; rationale and methods. Behavioral Science, 1957, 2, 216-221.

THE USAF ACADEMY HOW-TO-STUDY PROGRAM: PAST, PRESENT, AND FUTURE[1]

Mark Nataupsky, John Bermudez, Nita Huelf
and Valentin Tirman

United States Air Force Academy

This paper describes the Air Force Academy's How-to-Study program from three points of reference. First, there is a very brief look at the original scope and philosophy. Second, there is a detailed look at the program as it now functions. Third, there is a brief look at future goals and activities.

PAST

The Air Force Academy's How-to-Study program began as a very modest undertaking in 1972. It was an effort by a few individuals who wanted to do something extra to assist cadets who were in academic difficulty. The program was designed as a behavior modification effort which was based on the amount of time spent studying.

The key to any modification of behavior is an acute awareness of present behavior. With respect to academics, one key behavior is the amount of study time. It is very easy to fool yourself and to believe that you are actually spending a great deal of time studying, when most of the time is really spent in other activities. For example, time spent casually chatting during an evening is usually included in a cadet's concept of "study" time.

There were specific recommendations to help cadets keep track of actual time spent studying. It was first necessary to purchase an electric clock and to insert an in-line switch close to the clock. The cadet was asked to set the clock to 12 o'clock, to turn it on at the start of studying, and to turn it off whenever there was a break in studying. The clock then displayed the cumulative study time and the time was recorded on a log. When a different subject was studied, the clock was first reset to 12 o'clock in order to see easily the cumulative study time for that subject.

There were several benefits from the use of the clock if the above procedure was followed closely. First, the cadet received immediate, direct feedback on the amount of actual study time. This feedback often served as a shock to cadets who thought they had spent an entire evening studying when most of the time was really spent in other activities. Another benefit was that the cadet became acutely aware that studying had ceased when he or she turned off the clock. The increased awareness that there was a break in studying prompted the cadet to return quickly to the books. Although it required discipline to use the clock as directed, the use of it usually was very reinforcing for those cadets who followed the procedure.

There were 11 cadets who participated in the program during that first year. The cadets each used a clock, as described, to help keep accurate logs of actual time spent studying. The program seemed to help these cadets to improve their grades.

[1]This research was sponsored by the Frank J. Seiler Research Laboratory work unit 2303-F1-54.

PRECEDING PAGE BLANK-NOT FILMED

For the next several years the program received close scrutiny. Matched groups of cadets who were in or out of the program were compared. Throughout this series of tests of the effectiveness of the program, it appeared that the cadets did, in fact, benefit from participation.

In 1977, the Academy Board (a school policy-making body) decided to expand the role of the How-to-Study program. Cadets with marginal admission qualification test scores were automatically enrolled in the How-to-Study program. Any of these cadets who later felt that either they had completed or were not benefiting from the program, were free to drop the program at any time. Any other cadets, including those with high admission scores who wanted to increase their grades were also free to join the program. The program continued to function as a behavior modification program which was based on techniques to increase awareness about time spent in study.

PRESENT

The enrollment philosophy today remains the same as it was in 1977. However, the program has grown considerably larger. For example, in the Fall semester 1979, there were 689 cadets who actively participated in the program. There were 183 Air Force and Exchange Officers who volunteered their services to serve as How-to-Study counselors for those cadets.

Cadets are still urged to purchase a clock and to use it as a monitoring device. Part of the program remains a behavior modification approach which is based on study hours; however, the present program involves considerably more activity than just behavior modification. Specific guidance is provided to the counselors (Bermudez, Huelf, and Nataupsky, 1979) in the form of suggested session plans. All counselors receive detailed training on the use of the session plans. The following describes these plans in some detail.

A topic closely related to increasing study time is the need to plan one's activities. We try to persuade the cadets to plan out activities for a full week in advance. In addition to scheduling classes and study time, cadets schedule time to relax, go skiing, or do anything else that is not related to academics. For example, if a cadet has plans to go to a movie on Saturday night, we want to see that on the weekly schedule. Seeing that movie scheduled for Saturday night often helps a cadet to stick to the study schedule during the week because the movie has reinforcing qualities.

Although the students' plans do provide a basic outline of activities, we also recognize that the plans often will be violated. It is most important for the cadet to understand why the plan was not followed. Cadets who can understand why their plans were not followed are in a much better position to formulate a more realistic plan for activities next week.

In addition to the requirements of weekly planning and scheduling, long range plans are also necessary. These plans include some form of a milestone chart. For cadets, it is particularly important to break a task into smaller tasks and to set their own deadlines for these milestones. A term paper might be divided into tasks such as (1) topic selection, (2) literature search (3) draft writing, (4) final writing, and (5) typing. The deadlines for those milestones might well be in conflict with the requirements for a major examination. Early in the semester, it is relatively

easy to shift some of the milestones associated with the paper. By planning all of the major requirements at the start of the semester, the cadet can balance his or her workload more appropriately. A little foresighted planning might even save the cadet from having to stay up all night in order to finish a paper on time. The more term papers and examinations which are required during a particular semester, the more the cadet stands to gain from planning a milestone chart.

Cadets are shown how they can exercise stimulus control of their study behavior. It does not make sense to continue to try to study in an environment where there are many distractions. Nonetheless, many cadets are unaware of the need to change locations if there are too many distractions present. Furthermore, it helps the cadet to have one or two different areas which are routinely used only for studying. Arriving at those locations provides a ready mental set to study. Study breaks of any kind are taken in a different environmental location.

Most cadets fail to realize they can give themselves effective reinforcements for their own study behavior. These reinforcements can vary according to the magnitude of the task which has been accomplished. For example, completing mathematics homework problems before a certain time in the evening might be rewarded with a soft drink instead of water at the next break. Another self-reinforcement for appropriate behavior might also be some free time to talk with friends. In contrast, completing a term paper by a set date might be reinforced with the purchase of a record album. Study behavior which is followed by self-reinforcement tends to strengthen the frequency of future study behavior. Although the principle is simple, cadets must be made aware of the self-control they can exercise by arranging contingencies.

Textbook reading skills are addressed by the SQ3R method. SQ3R stands for Survey, Question, Read, Recite, Review. In this way, we try to break the habit of inattentively reading through a chapter, and then having little recall or understanding what was read.

The typical cadet has a separate notebook for each class. The typical cadet also starts writing notes at the left margin, keeps writing until the right margin, and then drops down a line. There are alternative approaches to taking notes. One major change is to leave a very narrow margin along the left edge and a 2-inch margin along the bottom of the page. A few key words such as "titration" or "id" are written in the left margin. The area at the bottom of the page is used for a summary of notes on the page. A variation on this format is appropriate when the class presentations closely follow the text. With this format, the area for detailed notes is divided into two columns. One column is used for textbook notes. The corresponding class notes are written in the other column.

There is also a suggested session plan which briefly mentions some test-taking strategies. Cadets are urged to make a review schedule to avoid the common practice of cramming for examinations. They are told to write down specific facts to help organize their responses to possible essay questions. In an objective test, it is very important to avoid spending too much time on any one item. Changing answers is usually not a good idea unless the person is very sure of the answer. Above all, it is critical to read carefully the directions for both essay and objective examinations.

In addition to these suggested session plans, the How-to-Study office has approximately 60 helpful handouts which address specific issues. Cadets and counselors can, and do, get additional help from personnel in the office.

FUTURE

The issue of possible learning disabilities among cadets is just beginning to be addressed at the Air Force Academy. Learning disabilities have been defined in many ways. We use this description as a reference for cadets who are having great difficulty in learning specific academic skills in spite of the fact that they have above average intelligence. In this definition, we exclude cadets who simply do not spend sufficient time studying. We also exclude from this definition cadets who have a relatively low IQ scores on the Wechsler Adult Intelligence Scale (WAIS) and cadets who have roughly equivalent scores across the WAIS sub-scales. The typical learning disabled student has a relatively high overall IQ and also has great variations or scatter in the component scores of the sub-scales of the WAIS.

One specific area of concern is that of auditory perceptual problems. Cadets who have difficulty in this area do not have any problems with the reception of auditory stimuli, i.e., there is nothing wrong with their hearing. There is, however, something wrong with the way they process the information they hear. Cadets who have auditory perceptual problems typically exhibit one or more of a particular set of behaviors.

For example, one possible symptom is a lack of auditory attention. Cadets with this problem have considerable difficulty following a lecture, and they have frequent daydreaming periods. Of course, virtually everyone has these experiences occasionally but not to the extent of the learning disabled.

Two related problem areas are auditory discrimination and sequencing. These problems are most apparent when the cadet attempts to repeat spoken language. In addition to substituting words that sound alike, the sequence of sounds may be reversed.

These are just a few examples of typical behavior patterns which may indicate auditory perceptual difficulties. A few cadets have been evaluated and diagnosed as having auditory perceptual difficulties. They are receiving specific remediation at Fitzsimmons Army Medical Center in Denver, Colorado. These examples also point to the need for more extensive screening of cadets in the future.

Diagnosis of other forms of learning disabilities has turned out to be a difficult problem. A trained diagnostician is needed to interpret a battery of intelligence and other tests. We hope to have the services of a learning disabilities diagnostician during the 1980-81 academic year.

Relatively little is known about diagnosing and remediating adult learning disabilities. The Academy's How-to-Study program is dedicated to developing a program which can help cadets who have this difficulty.

REFERENCE

Bermudez, J.M., Huelf, N., and Nataupsky, M. The USAF Academy How-to-Study Program: A Handbook for Volunteer Counselors. FJSRL-TR-79-0003, ADA 071241. United States Air Force Academy, CO: Frank J. Seiler Research Laboratory, April 1979.

Individual Choice, Personal Affect, and Personal Adjustment of Women Who Have Resigned From West Point

Major Jerome Adams COL Howard T. Prince II Dr. Francine Hall
U.S. Military Academy West Point, NY

Abstract

This study examines the problem do women leave largely as a function of personal characteristics as opposed to external influences within the environment. A follow-up mail out survey was sent to all women of the Classes of 1980, 1981, and 1982 who resigned from West Point. The findings were interpreted in terms of two issues (1) a matter of better candidate selection and (2) implications for modifying aspects of the organizational climate and training.

INTRODUCTION

Like all institutions of higher learning, the United States Military Academy would like to minimize cadet attrition while maintaining high institutional standards. Previous research on attrition at West Point has found that the largest percentage of resignees leave because of motivational reasons. However, a number of organizational characteristics may contribute to a cadet's decision to leave (Butler, 1974). In 1976 a report by the General Accounting Office identified a number of factors relating to attrition at the five federal service academies (GAO FPCD-72-12). For example, the report found that the percentage of attrition which can be attributed to the Academy environment, as opposed to external factors or student characteristics, increases significantly over time.

With the passage of Public Law 94-106, women were admitted to the service academies in the summer of 1976. Reports on the study of the admission of women point to a higher attrition rate for females compared to male cadets (Vitters & Kinzer, 1977; Durning, 1978).

The purpose of this paper is to examine: (1) the reasons why women have attritted from West Point, (2) the personal feelings or affect of women who have left, and (3) the personal adjustment women have made after their resignations.

Theoretical Issues: Within the literature of the behavioral sciences, there is no body of knowledge under the rubric, Theories of Attrition. Quite aptly, one may point out that attrition is not an abstract concept which, once

conventionally defined, must be operationalized and measured. Rather, attrition is a descrete action. It is a behavior or the result of a choice an individual selects to end commitment. Within the academic area, a number of authors have contributed to our knowledge and understanding of the reasons for attrition through their studies about the concept of commitment. Kanter (1972) defined commitment in terms of an individual's willingness to devote energy and loyalty to an organization. Porter, Steers, Mowday, & Boulian (1974) defined commitment as the individual's belief in and acceptance of the organization's goals and values. In a recent review of the literature on organizational commitment, Hall (1979) concludes that the concept of commitment has been defined based upon a combination of attitudes and behavioral intentions.

In terms of operational measurement, Hrebiniak & Alutto (1972) have used a simple and straightforward measure of commitment -- the individual's degree of willingness to leave the organization. Researchers in the military concerned about measuring attrition have adopted the straightforward measurement of commitment. For example, at West Point the most effective measure used to predict attrition of cadets is a simple Academy Graduation Commitment Gradient, AGCOG (Priest, 1977).

A problem pointed out by Hall (1979) with the behavioral intent measure of commitment is that forces other than commitment can affect commitment choice. For example, romantic interpersonal attractions between coeds at West Point may cause conflicts in the person's choice to stay or resign. More importantly, factors within the organization's environment, e.g. non acceptance by male cadets, may have influenced the women's choice to leave. Finally there has been a concern on the part of officials at West Point about the validity of cadet responses during exit interviews. More specifically, there is a suspicion that once a cadet decides to leave, the cadet may formulate a response reflecting what the cadet thinks the officials want to hear rather than how the cadet may truly feel (Adams, 1979).

METHOD

Subjects. The subjects in this study included all of the women who were ever formally admitted into West Point in the Classes of 1980, 1981 and 1982, who subsequently resigned. The incoming Class of 1983 was admitted during the conduct of the study, and they were not included in the sample.

Procedure. A follow-up mail out survey was sent to all women in the Classes of 1980, 1981 and 1982 who resigned. The total population was 110. Of that group 68 completed the answer sheets accurately and an additional 4 elected to return long personal letters in lieu of the answer sheets. There was also one personal letter response from a parent explaining the reasons for the daughter's resignation. Including the personal narrative responses 66% of the women contacted responded. In addition, 4 questionnaire packets were returned by the postal service because of change of address.

Analyses. The responses given by the former female cadets were analyzed to answer the following questions:

- The reasons they gave for leaving during exit interviews versus how they look back on the reasons now.
- How they felt personally about their former Academy experiences and their eventual decision to leave.
- How significant others have reacted to their decision to leave.
- What their current education and marital status are.

Results. In the first two items when cadets were asked "why you actually resigned" and "what was the reason you gave," differences were found. These discrepancies support the earlier suspicion that information obtained during the exit interview often may not accurately portray the true reasons why women resign.

Regarding personal feelings about their Academy experiences and their decision to leave, the majority of the women (67%) reported favorable feelings at departure and now. There were 12% who reported unfavorable feelings at departure and now. Also, 66% of the women reported that they felt that they made the right choice to leave. However, there were 19% who stated that they were uncertain as to whether resignation was the right choice. A majority of women 69% stated that if asked they would recommend that other women apply for admission to West Point. There were 24% who stated that they would not recommend that other women apply.

In the analyses of the importance of support from significant others to the decision to leave, 99% stated that they found support from either the family or from friends after they left. However, the women were also asked questions about support they could obtain from people within West Point prior to leaving. The majority of the former female cadets report that they did not feel comfortable to discuss their doubts about staying at West Point with the Tactical Officer or with members of the Cadet Counseling Center.

In response to the present education status 73% stated that they were full time students in college. Also, 85% of the women reported that they were still single.

DISCUSSION

The results reported in this study indicates that women provide some answers during exit interviews which differ when asked months or years later to reflect on those same choices. These data also show that there are many factors which influence a female cadet's decision to resign. This confirms the position stated earlier that forces other than commitment do affect commitment choice.

The finding that the majority of women who resigned have positive personal feelings about their academy experiences and that they were able to obtain some favorable support from either family or friends is encouraging. This is the first systematic feedback which West Point officials have received regarding the positive feelings women have about the arduous and demanding experience of Academy life. It is also encouraging to note that most women, if asked, would recommend that other women seek admission.

The finding that the women were uncomfortable to discuss feelings to resign with officials in the formal support system is not new. Often cadets especially new cadets, view their TAC's in antithical roles one as a counselor, the other as a disciplinarian. There are also stereotypes within the Corps of Cadets that suggest a person is weak if it is known that the cadet wants to go to the Counselling Center to discuss a problem. Such labeling may contribute to the reluctance of the women to freely go to the Counselling Center to discuss doubts about organizational commitment. Based upon the respondents, the majority of women are enrolled in college and still are not married.

The issues and concerns about why women have left West Point will continue to be studied. There are many questions still unanswered. For example, the majority of women answering this survey reported that their physical performance did not strongly influence their decision to leave. Yet, overwhelmingly the women also stated that their physical performance was a major criteria for acceptance by their male peers. This seems to support ongoing research which has linked female cadet leadership evaluations with how they perform in physical performance tasks.

PRECEDING PAGE BLANK-NOT FILMED

REFERENCES

Adams, J. Report of the Admission of Women to the U.S. Military Academy (Project Athena III) West Point, NY: AG Printing Office, June, 1979.

Butler, R. P. Attrition and Management Style (Project 188). Unpublished paper West Point, NY: Office of Institutional Research, April, 1974.

Durning, K. P. Women at the Naval Academy: An Attitude Survey. Armed Forces and Society, 1978, 4, 585.

Hall, D. T. Organizational Commitment: Theory Research and Measurement. Unpublished paper Evenston, Ill: Northwestern U., April, 1979.

Hrebiniak, L. G. & Alutto, J. A. Personal and Role Related Factors in the Development of Organizational Commitment. Administrative Sciences Quarterly, 1972, 17, 555-573.

Kanter, R. M. Commitment and Community. Cambridge, Mass: Harvard U. Press, 1972.

Porter, L. W., Steers, R. M., Mowday, R. T. & Boulian, P. V. Organizational Commitment, Job Satisfaction, and Turnover Among Psychiatric Technicians. Journal of Applied Psychology, 1974, 59, 603-609.

Priest, R. F. Understanding the Commitment of Cadets to the Academy and to a Military Career (Project 203). Unpublisted Paper West Point, NY: Office of Institutional Research, 1977.

Student Attrition at the Five Federal Service Academies (U. S. General Accounting Office FPCD-76-12). Washington, D. C.: Comptroller General of the United States, March, 1976.

Vitters, A. G. & Kinzer, N. S. Report of the Admission of Women to the U.S. Military Academy (Project Athena I). West Point, NY: AG Printing Office, September, 1977.

Male and Female Differences in Army Career Branch Selections at West Point

Cadet Janice Calhoon and Cadet Bret Dalton
U.S. Military Academy, West Point, NY

Abstract

This study proposes to answer the question, "Why do cadets choose the branches they do?" Three classes of independent variables: personal characteristics, environmental experiences, and expected satisfaction were used to determine their influence on branch selection. The implications of the results for Academy training programs are discussed.

INTRODUCTION

Choosing a branch is a major decision which will influence a cadet's future in the military after graduation. Each cadet chooses a branch or specialty during January of the senior year. Branch selections are made toward the end of a cadet's experience with the hope that each cadet is informed well enough to make an intelligent and satisfying decision. This study attempts to discover which factors leading to branch/specialty choice are important to the cadet. The study examined the influence of personality factors, environmental experiences and expected satisfaction on a cadet's branch or specialty choice. This study will consider theoretical and practical issues.

Theory states that learning is accomplished vicariously through role models. One question we hope to answer in this study is how much do influential others (role models) have on a cadet's eventual branch/specialty selection. Practically speaking, we also expect that a cadet's personal values and career aspirations may be quite significant in the selection process. That is, cadets are concerned about both personal values as well as professional aspirations when they choose a branch or specialty. Cadet summer training programs have a varying affect on the ultimate decision, depending upon which programs are attended. Another information source we will consider is the Department of Military Instruction (DMI) sponsored branch/specialty orientation. Another parameter which influences branch/specialty choice is the Performance Ranking of all seniors (GOM). Finally, a number of experienced officers have reported the importance of satisfaction as an influence on branch/specialty selection.

This study, to the best of our knowledge, is the first of its kind. Hopefully, similar studies will follow.

Problem: The question we propose to answer is, "Why do cadets choose the branches they do?" An individual bases his/her branch/specialty selection on the information available. From the time an individual arrives at West Point, the cadet continually receives information that will eventually affect this decision. This information comes from numerous sources which vary in degree of influence. For example, influential others are perhaps the most frequently encountered source of information. The cadet's professors, tactical officers, and mentors at the Academy are for the most part commissioned officers in the Army. These officers serve as role models. Research in other organizational settings have found that Social Learning Theory is applicable to influencing the behavior of others (Bandura, 1977).

A second factor influencing the final choice is the individual cadet's personal values. For example, upon entering the Academy, 90% of male cadets and 73% of female cadets expressed a desire for a career and a family (Adams, 1979). Cadets' career aspirations are considered to be equally weighted with personal values as a major influence on these decisions. Many cadets choose a branch/specialty which teaches a skill or gives experience which is marketable in the civilian job world. Job satisfaction can increase quality of performance and productivity in any occupation (Hackman, Oldham, Janson, and Purdy, 1975). Thus, maximization of satisfaction should be an unwritten desire in this selection process.

Another factor which influences selection of branch/specialty is education opportunities available to cadets in the summer training programs. These programs are mandatory for all cadets and are sponsored by the Department of Military Instruction. For example, Cadet Troop Leader Training is one of the programs which provides cadets with their first opportunity to serve as a junior officer in a Regular Army Unit. The cadet can experience first hand what it may be like to serve in a particular branch or specialty. In another summer training program upperclass cadets are afforded the opportunity to instruct and train freshmen in Cadet Basic Training (CBT) and sophomores in Cadet Field Training (CFT). Seniors also have a chance to work with branch specialized training units while serving on training committees at CFT. Exposure in these training roles provides the cadet with a limited experiential base to judge how well the cadet may like duties in this specialty. Other available training programs include: Ranger, Airborne, Northern Warfare, Jungle, Air Assault, and Flight.

A branch/specialty orientation is offered by DMI as

another source of information to assist cadets in branch selection. This orientation consists of a series of lectures, seminars, publications, and individual branch/specialty representatives.

Another factor influencing the cadets' choice is his/her Performance Ranking (GOM). This ranking is used at West Point in part to provide equity in the choice process. Cadets who perform well are rewarded with their first choice of branch/specialty. However, choice is also limited by quotas. There are specific quotas that must be filled in each branch/specialty.

The last factor to be examined in this study is the influence of expected satisfaction. Two ways of looking at this issue are (1) cadets' overall satisfaction with career field chosen, and (2) the expected satisfaction with the first assignment.

Many other inputs may influence an individual cadet's decision, but the above stated inputs are generally the most significant. Therefore, these will be the considerations studied.

The purpose of this study is to determine what influence each of the factors has in branch/specialty selection.

METHOD

Upon the selection of branch/specialty by members of the Class of 1980 at the United States Military Academy, a field study was conducted using a locally developed questionnaire. The instrument included a section on personal attributes, environmental experiences, and expected satisfaction. The purpose of the study was to investigate the important reasons why men and women choose the career branches/specialties they do. The data gathering methods manipulated three classes of independent variables which influence branch/specialty selection.

The first class of variables included personal characteristics (e.g. sex, life style preference, and other demographic information). Life style preference, was operationalized by a series of questions relating to individual values of marriage, career and family. Other demographic information included sex, age, and prior service.

The second class of variables included environmental experiences during cadet development (e.g. summer training, influential others, and Academy sponsored branch/specialty education).

The third class of independent variables includes expected satisfaction (e.g. satisfaction with profession and satisfaction with first assignment).

The criteria assessed were the actual branch/specialty choice desired (e.g. Combat Arms, Combat Support Arms and Combat Service Support Arms). The branches listed in Combat Arms for Academy purposes are Infantry, Armor, Field Artillery, Air Defense Artillery, Aviation and Engineers. The branches listed in Combat Support Arms are Signal Corps, Military Police, Military Intelligence, and Chemical Corps. The branches listed in Combat Service Support Arms are Transportation Corps, Ordinance, Quartermaster Corps, Adjutant General's Corps and Finance.

Sample: Sixty cadets from each of the categories (Combat Arms, Combat Support and Combat Service Support) were sampled. Within those categories all of the women were sampled, i.e. in Combat Arms twenty-seven women were sampled and thirty-three men were sampled for a total of sixty in that category. A total of 180 cadets participated in this study.

Procedure: Individuals were administered a forty item questionnaire during their branch/specialty orientation course. Individuals were selected by name. The answers were coded on standardized answer sheets. The questionnaires were given in a classroom setting, and the answer sheets were collected immediately at the end of the administration.

Results: The results of the study were interpreted to allow senior Academy officials to gain a better understanding of what the influence of personal characteristics, environmental experiences, and expected satisfaction have on cadets as they make their final choice of branch or specialty. The data were analyzed using a series of multiple regressions to determine the best predictors of subsequent branch/specialty selection from these factors.

DISCUSSION

In any decision making process it is valuable to examine the relationship between personal characteristics, environmental experiences, and expected satisfaction with the eventual outcome. In this study we were trying to provide answers relating to how much these factors had on cadets' choice of branch or selection. This is the first study to look, in a systematic manner, at what variables are used and how a cadet weighs these variables in his/her decision. This is the first real sensing of what effect co-education has had on a cadet's branch/specialty preference. This is the first year life style preferences have played a more significant role with respect to dual career planning. Senior Academy

officials can use the findings as input to reassess how effective information sources are in the present system.

Many of these factors have never been systematically studied. Because of the transition the Academy is presently undergoing, new factors are becoming paramount and their effects must be considered for future policymaking.

REFERENCES

Adams, Jerome. Project Athena III. West Point, N.Y.: AG Printing Office, 1979.

Bandura, Albert. Social Learning Theory. Englewood Cliffs, N.J.: Prentice-Hall, 1977.

Hackman, J. R., et al. "A New Strategy for Job Enrichment," California Management Review, 27 (Summer 1975), 62.

Cadet Interest in Flying Specific Aircraft

Jefferson M. Koonce
United States Air Force Academy, Colorado

Abstract

Expression of a belief that Academy Cadets would definitely prefer to fly Fighter-Attack-Reconnaissance (FAR) type aircraft instead of Transport-Tanker-Bomber (TTB) type aircraft was tested by a survey of the pilot qualified first class cadets in the Class of 1979. The results support the hypothesis of preference for FAR aircraft for immediate assignment, but not in terms of long term or career assignments.

The Air Force's present undergraduate pilot training (UPT) program is designed to produce universal pilots potentially capable of being assigned into any weapon system of the United States Air Force. Such a broad based development of a large spectrum of flying skills is not economically practicable in this day and age when flight training is more expensive than ever. These 'universal' pilots are taught some flying skills that may not be used by them in their individual assignments following pilot training, while other skills required in their first assignments may receive only limited training.

Consideration of establishing a Specialized Undergraduate Pilot Training (SUPT) system has been made with the intent of maximizing the training effectiveness, being as cost effective as possible, and producing pilots that are more qualified for their first assignments following pilot training. One concept is that at the end of the primary phase of pilot training a decision is made as to whether the pilot trainee should continue pilot training in a track designed for fighter-attack-reconnaissance (FAR) aircraft or one designed for transport-tanker-bomber (TTB) aircraft. At that time the student pilot would have more awareness about the profession of being a pilot, his instructor would have had an opportunity to observe the student's flying abilities, and the Air Force's projected pilot needs (TTB and FAR) would be more definitive. The student would be motivated to perform his best during primary training, consideration of his desires would not adversely affect his morale, and the retention of graduating pilots would be enhanced by placing them in the most suitable track commensurate with their demonstrated abilities and desires. Invariably there will be occasions in which the student's desires and abilities do not coincide, and his instructor who has observed his performance would have to counsel him about accepting that which is best in terms of his ability and the Air Force's needs.

In considering personal preference as a factor in dual track assignment, the belief that the FAR track would be overwhelmingly desired in lieu of TTB had been frequently discussed. Because there was considerable disagreement in the officers' perceptions of the Cadets' desires for flying assignments, a brief questionnaire was developed to actually determine the cadets' interests in one track or the other. Only five questions were asked. The first determined the cadet's current medical status: pilot, pilot waiverable, navigator, navigator waiverable, and non-rated. The

PRECEDING PAGE BLANK-NOT FILMED

second and third questions were aimed at determining which aircraft they would like (question 2) or dislike (question 3) most to be assigned to and fly at that particular moment in time. The cadets were asked to rate their top choices 1, 2, and 3 and the least desirable three as 7, 8, and 9. Question 4 was aimed at determining which aircraft they would realistically expect to be assigned to upon completion of pilot training, and question 5 asked what aircraft they would prefer on a long-term or career basis.

Results

The responses were made on digitex answer sheets and machine scored. The total number of responses between questions and within the multiple answer questions varied because of errors in coding responses and because some cadets did not answer all of the questions. The responses to individual aircraft were condensed by combining them into the categories of Strategic, Attach, Helicopter, Trainer, Cargo, and Fighter, and subsequently further reduced to TTB, FAR, and Other (helicopter and trainers).

Question 1. "At this specific time in your career, which of the following aircraft would you MOST like to be assigned to and fly?

	Strategic	Cargo	Attack	Fighter	Helo	Trainer
1st Choice	17	178	34	273	40	21
2nd Choice	24	188	58	229	44	39
3rd Choice	50	188	83	179	30	48

	TTB	FAR	Other
1st Choice	195 (34.6%)	307 (54.5%)	61 (10.8%)
2nd Choice	212 (36.4%)	287 (49.3%)	83 (14.3%)
3rd Choice	238 (41.2%)	262 (45.3%)	78 (13.5%)

Referring to the above table, a statistical test of the differences in proportions indicated that desire for FAR at the present time is significantly greater than for TTB (p 0.001) and desire for helicopter and trainer (Other) was significantly less than for either TTB or FAR (p 0.001). The results are similarly significant when considering the cadets' second choices. But, on their third choice, Other was significantly less than TTB or FAR while there was no signficant difference between TTB and FAR.

Question 2. "At this specific time in your career, which of the following aircraft would you LEAST like to be assigned to and fly?

	Strategic	Cargo	Attack	Fighter	Helo	Trainer
9th Choice	383	23	28	13	14	63
8th Choice	218	46	97	23	25	124
7th Choice	151	101	111	23	35	120

	TTB	FAR	Other
9th Choice	406 (77.5%)	41 (7.8%)	77 (14.7%)
8th Choice	264 (49.5%)	120 (22.5%)	149 (28.0%)

With regard to the least desirable aircraft to be assignes to and fly at the present time, the TTB are significantly less desirable than the FAR or Other at the 9th, 8th, and 7th levels of choice (all $p < 0.001$).

Question 4. "Realizing that not all UPT graduates get their first choice, which one of the following aircraft do you realistically expect to be assigned to upon completion of UPT?"

Strategic	Cargo	Attack	Fighter	Helo	Trainer
49	191	43	193	29	33

TTB	FAR	Other
240 (44.4%)	236 (44.0%)	62 (11.6%)

Question 5. "After completing your first rated assignment following UPT, which one of the following aircraft would you like to be assigned to on a long-term or career basis?"

Strategic	Cargo	Attack	Fighter	Helo	Trainer
10	231	30	201	36	25

TTB	FAR	Other
241 (45.0%)	231 (43.5%)	61 (11.5%)

The cadets' expectations of TTB or FAR assignments following completion of UPT are not significantly different, and fewer cadets anticipate being assigned to Other than to TTB or FAR ($p < 0.001$).

With respect to aircraft desired on a long-term or career basis, cadets desire TTB or FAR over the Other ($p < 0.001$), but there was not a significant difference between their desire for TTB and FAR.

Discussion and Conclusions

The hypothesis that cadets would prefer to fly FAR aircraft was supported in terms of immediate desires, but in terms of long-range or career desires it was not. At the time of the survey, cadets would prefer to fly the F-15/16 aircraft most (N=258) followed by the C-141 (N=104). In terms of least desired, only one aircraft clearly stood out, the B-52 (N=357). In the category of next to least desirable there was a sizeable increase in the number who identified the KC-135 (N=122), FB-111 (N=44), and the T-41 (N=91). The lack of desire to fly T-41s might be, in part, a reaction to the recent T-41 training experienced by the cadets and also a desire to fly the "real" Air Force aircraft. While the tactical aircraft (F-15/16) might always appear exciting and desirable to the young future aviators, one might take care in developing information packages with regards to the Air Force's strategic aircraft so that they are not portrayed as totally undesirable. There will be a need for strategic aircraft pilots in the Air Force for some time to come, and a good information program about that career area would assist in insuring a supply of pilots desiring to fly those aircraft.

Despite their current desires for particular aircraft, the Academy cadets exhibited rather realistic expectations with regard to the types of aircraft they would most probably be assigned to upon completion of pilot training. A point of concern is what affects upon morale and/or retention

is caused by the dissonance between the aircraft assignments desired and those expected. This difference seems to be somewhat resolved in the cadets' expressed desire for certain long-term or career assignments. Looking at their responses to questions 4 and 5 in light of their desires, reflected in questions 2 and 3, one sees that there is a sizeable increase in the numbers desiring cargo aircraft with a commensurate decrease in strategic aircraft. In the same fashion, the number of anticipated F-4 assignments after pilot training show a decrease on the long-term question while the F-15/16 aircraft have a sizeable increase. However, in the category of cargo aircraft there is an overall increase, especially for the C-5 and C-9 aircraft. Much of this increase might be attributable to the young pilots planning to develop alternative career options or merely the shift from strategic to cargo aircraft.

In general, the resulta show a current preference for FAR over TTB aircraft but no significant difference between FAR and TTB with respect to career goals. Although there might not be a large difference between the proportions of cadets desiring TTB and FAR assignments, there should be considerable concern about the type of FAR and TTB aircraft desired or not desired. The rather considerable lack of desire for strategic aircraft assignments suggests the need for a carefully planned information program with respect to career in strategic aircraft to meet the future needs of the Air Force.

A problem which should be further researched is just what are the sources of information from which cadets develop their attitudes towards specific aircraft and serving in the various commands.

STRESS AND ITS SOURCES IN A MILITARY TRAINING ENVIRONMENT

Cadet Douglas McCord and Colonel Howard T. Prince II
United States Military Academy
West Point, New York 10996

Abstract

This paper reports the results of a project designed to examine the impact of specific environmental stressors on cadets during summer training at West Point.

Introduction

One dominant characteristic of the Cadet Basic Training (C.B.T.) experienced by New Cadets during their first summer at the United States Military Academy is the stress they encounter. The responses of New Cadets to the stress in this training environment, and some of the possible sources of this stress are the central issues in this study. Stress may be viewed as a set of stimuli that evoke performance hampering emotional responses (Zimbardo, 1977). One of the primary theories of emotions focuses on the Psychological Responses to external stimuli. In a series of studies on emotions (Schachter and Singer, 1962; Schachter, 1971), subjects experienced both chemically-induced physiological arousal and an external stimulus for cognitive appraisal in the form of the behavior demonstrated by another individual in the same room. The subjects expressed different Psychological Responses to the same physical stimuli based on their cognitive appraisal of their surroundings (the other's behavior). Some of the negative Psychological Responses associated with stress that would bear examination are Anxiety, Hostility, and Depression (Zuckerman and Lubin, 1965).

Another theory of emotions (Arnold, 1960) suggests that the physiological arousal component of an emotion does not precede cognitive processes such as appraisal, but rather the affective response and physiological arousal are outcomes of the cognitive appraisal of the situation. Thus, Physiological Responses to stress were also examined.

Any environmental stimulus that causes undesirable or excessive psychological or physiological responses would be a Stressor. The primary sources of stress within the C.B.T. environment fall under the following categories: the Standards of Performance set for the New Cadet by the institution; the agents of the institution that enforce these demands, Formal Authority; and the individuals supporting the New Cadet and encouraging cadet success, Support System. This concept suggests that a beneficial study might be made of the relationships between the Psychological and Physiological

Responses to the stress experienced by New Cadets and the external stimuli within C.B.T. that might be associated with this stress.

Another line of inquiry suggested by this research is whether differences exist between the responses of males and females to stress in summer training. Previous research (Prince, Leister & Deller, 1979) showed that, except for the pioneering class (of 1980), no differences could be detected between the psychological responses to stress of men and women during C.B.T. Concern was expressed by the Academy's administration as to whether training of a different nature could lead to such a difference. During their second summer at West Point, cadets undergo military training at Camp Buckner. The training there is more demanding physically than the training in C.B.T. The research supported by Project Athena (Adams, 1979) reports that female cadets had expressed a concern over the expected difficulty of the physical demands at Camp Buckner. Thus previous research suggests that differences might be found between the psychological responses to stress exhibited by men and women in a more physically demanding training environment such as Camp Buckner. The most strenuous portion of the training at Camp Buckner is the "Recondo" training in patrolling and survival. A comparison of the degrees of stress (as indicated by overall Psychological Responses) experienced by cadets at the start of the Camp Buckner training (baseline), immediately before the "Recondo" experience, and immediately after "Recondo", would then be of interest. This comparison would indicate how cadets react psychologically to stress in various situations, such as the anticipation of a demanding experience versus the recovery from that experience.

Method

The data for this study of stress were collected during the Cadet Basic Training at West Point for the Class of 1983. The study consisted of the cross-sectional administration of a questionnaire. Two to three New Cadets from each squad in training were randomly chosen as subjects. A total of 240 subjects were surveyed on the fifth day of training and 192 were surveyed on the forty-first day. The Psychological Responses to stress of Anxiety, Hostility, and Depression were measured by the Multiple Affect Adjective Check List (MAACL) (Zuckerman and Lubin, 1965). Physiological Responses were measured by questions as to the number of times a subject experienced certain physical symptoms related to stress. The Stressors, Standards of Performance, Formal Authority, and Support System were measured by questions as to the degrees to which the subjects perceived stress from certain people and events. Data were also collected for the Class of 1982 at

Camp Buckner. Psychological responses were measured by the (MAACL) questionnaire on the first day of summer training (baseline measurement), approximately two hours before going out on "Recondo Training," and approximately one hour before the end of the "Recondo" experience, before they knew whether or not they had successfully completed this training.

Results and Discussion

The results of this study are given in Table 1. Overall, Formal Authority was the best predictor of both psycholcgical and physiological reactions to stress. The influence of Standards of Performance on cadet psychological response to Hostility and Depression increased considerably from the beginning of C.B.T. to the end. This increase could perhaps have resulted from a clearer understanding by the New Cadets of what those standards were, or an increasing concern with meeting those standards. The effect on Anxiety was fairly constant.

The strength of relationship of Formal Authority on all criteria increased from the beginning of CBT to the end. However, at each point in time measured, there was an inverse relationship between Formal Authority and each of the criteria except Hostility. One possible explanation for the direction of these relationships is that the responses to stress other than Hostility, were caused largely by an uncertainty regarding the standards set by the institution. The Formal Authority was perhaps viewed by the new cadets as the primary means of better understanding those standards; thus, Formal Authority could have the effect of decreasing the uncertainty about what was to happen next, and is reflected in the related decrease in the Cadets responses to stress.

The only Response variable that showed a positive correlation with Formal Authority at the end of C.B.T. was that of Hostility. One possible reason for this is that although the Formal Authority helped clearly define the standards, these standards were perhaps considered by the New Cadets to be unattainable, unrealistic, irrelevent, or even inappropriate, thus causing frustration and resentment. This information and resentment could then have been manifested in reported feelings of Hostility. Another possible source of frustration and resentment manifested by hostility might have been a perception by the New Cadets that the time they were given was insufficient to allow them to increase their performance in order to meet the standards clarified by the members in Formal Authority.

An interesting proposition examined in this analysis was the influence of Support System as a possible source of stress for the New Cadet during C.B.T. From one perspective, the Support System of the family, friends and classmates of the New Cadet might be viewed as a source of moral support and be expected to help reduce the cadet's reactions to stress. On the other hand, the expectations of Support System for the individual to perform well and succeed might increase the stress that he feels. The results suggest that the Support System does not act as a Stressor, but may act to reduce the stress experienced to some degree.

The second major hypothesis focused on whether differences existed between the psychological responses to stress of men and women during the training at Camp Buckner. The results are given at Table 2. The data reveal that women entered the training with a higher level of Psychological Responses to stress than did the men. The measurement taken for men and women immediately before the "Recondo" experience was only slightly higher than the baseline measure, suggesting that the anticipation of "Recondo" had a minimal effect on Responses to stress. However, the measurement taken immediately after "Recondo" was much higher than the baseline for both men and women. This evidence suggests that having experienced a very demanding situation could have much more effect on psychological responses than the anticipation of that situation. Furthermore, the differences in the levels of Responses to stress for men and women suggest that some strategies for coping with this stress might be more effective for one sex or the other.

Conclusions

The implications of the results of these studies suggest two possible approaches to the management of stress. (1) A clearer, more comprehensive understanding of the causes of responses to stress experienced by cadets in summer training could make possible a decrease in dysfunctional stress through control or elimination of its sources. The goal in this case would be to enhance cadet performance by removing unproductive stressors. For example, responses to stress experienced by New Cadets at meals, partly resulting from requirements to recite memorized information, were significantly reduced after the requirements to recite information were eliminated. (2) Perhaps a more feasible approach to stress management is to seek, through the clearer understanding of the causes of stress, more effective strategies for coping with the stress. These coping strategies would be used to minimize the adverse

effects of the stress experienced, because sources of stress obviously cannot always be avoided. One such strategy that has proven useful in coping with stress is the method of "cognitive Rehearsal". This method involves the imagination of a stressful situation before it occurs as a form of emotional "preparation" (Folkins, Lawson, Opton, and Lazarus, 1968). Another strategy used in the study above was relaxation training. The difference in the levels of responses to stress for men and women in the Camp Buckner training environment suggests that different coping strategies might be used more effectively for women or for men. The strategies utilized should focus on the specific environment stressors involved. For example women expressed a greater concern over the expected physical demands of Camp Bucnker than did men. Thus, the coping strategy employed for women should address their concern about physical demands as a possible source of stress. On the other hand men may be more concerned about the possibility of failing the "Recondo " training and thereby losing the esteem from their classmates. Thus, the coping strategy employed for men should address their concerns about self image and esteem.

The relationships between the stressors and the responses to stress studied here are potentially useful considerations for senior Academy officials when reevaluating training and development programs at West Point.

References

Adams, J. Report on the Admission of Women to the United States Military Academy (Project Athena III) West Point, NY: AG Printing Office, June 1979.

Braun, J. and Linder, D. E. Psychology Today: An Introduction New York: Random House, 1975.

Prince, H. T., Leister, A. & Deller, W. J. "Stress in Cadet Basic Training" West Point, NY: Unpublished Report, USMA, February 1979.

Zimbardo, P. G. Psychology and Life. Glenview, Illinois: Scott, Foresman, 1977.

TABLE 1

R^2 STATISTICS FOR CADET BASIC TRAINING

	Fifth Day	Forty-first Day
HOSTILITY		
STANDARDS OF PERFORMANCE	.01110	.02822
FORMAL AUTHORITY	.00433a	.02049
SUPPORT SYSTEM	.00100a	.00248a
DEPRESSION		
STANDARDS OF PERFORMANCE	.01866	.03356
FORMAL AUTHORITY	.00213	.03934 a
SUPPORT SYSTEM	.00092	.00003
ANXIETY		
STANDARDS OF PERFORMANCE	.03645	.03489
FORMAL AUTHORITY	.00038	.06334 a
SUPPORT SYSTEM	.00197	.00792 a
PHYSIOLOGICAL		
STANDARDS OF PERFORMANCE	.00173	.01937
FORMAL AUTHORITY	.00337a	.02588 a
SUPPORT SYSTEM	.00268a	.00060

aIndicates inverse relationship

TABLE 2

MEANS FOR COMPOSITE STRESS RESPONSES FOR CAMP BUCKNER

FEMALES

BASELINE	BEFORE	AFTER
37.785	39.697	50.187

MALES

BASELINE	BEFORE	AFTER
28.986	32.995	44.493

PRECEDING PAGE BLANK-NOT FILMED

An Examination of the Relationship Between Supervisory Feedback and Job Performance Among Specific Air Force Ethnic Groups

Fulton D. Hayes, Major, USAF
Air Command and Staff College
David A. Wilkerson, Major, USAF
Air Force Leadership and Management Development Center
Maxwell Air Force Base, Alabama

Abstract

This study focuses on the relationship between supervisory feedback, one aspect of organizational communications, and employee productivity. The effects of supervisory feedback on the perceived quantity and quality of employee output were observed, and perceptions of male and female members of the following five ethnic groups were evaluated: American Indians or Alaskan Natives; Asians or Pacific Islanders; Blacks, not of Hispanic origin; Hispanics; and Whites, not of Hispanic origin.

Increased employee productivity has long been established as a primary goal of managers at every level in virtually every conceivable organizational setting. This is especially true in today's environment of scarce resources. Writing in Personnel Journal, Peter Kirby, a Vice President for Human Resources in the General Host Corporation, described productivity (or the lack thereof) in both manufacturing and service industries as a critical problem (1977). In a recent study sponsored by the Air Force Human Resources Laboratory (Pritchard, Montagno & Moore, 1978), Purdue Research Foundation analysts identified the problem of enhancing productivity as a central concern in today's Air Force. While there are undoubtedly a number of reasons for decreasing employee productivity, it is becoming increasingly more evident that one of the major culprits is ineffective organizational communication. This fact has been substantiated by personnel of the Air Force Leadership and Management Development Center (LMDC) (1978) and by non-Air Force management researchers, teachers and practitioners as well. It has been determined that, for most situations, the effectiveness and efficiency of willing, trained workers is directly related to the quality, quantity and timeliness of the specific information communicated to them (Glover, 1958). If management entails getting things done through people, communications is the essence of it; for without effective communications no one would know what they were supposed to be doing. Nor would there be any basis for answering questions, solving problems, obtaining feedback or measuring results (Hodgetts, 1975).

Realizing that feedback is indeed a multi-dimensional phenomenon, the following definition of the concept was used for this analysis: Feedback is information regarding some aspect of performance relayed from the immediate supervisor to a subordinate. The analysis was designed to provide answers to a number of managerial questions: Do

PRECEDING PAGE BLANK-NOT FILMED

Air Force supervisors provide different kinds of feedback to "employees" of different ethnic groups? Do some groups get more negative or positive feedback than others? Do males and females receive essentially the same quality and quantity of feedback from supervisors? What is the impact of the various dimensions of feedback on the performance of Air Force members?

Approach and Methodology

The Survey

In this examination of the relationship between supervisory feedback and job performance, data obtained through the administration of the Organizational Assessment Package (OAP) was used. The OAP, a survey instrument used to assess leadership/management strengths and weaknesses of Air Force organizations, was jointly developed by the AF Leadership and Management Development Center (LMDC) and the AF Human Resources Laboratory (HRL). The "package" from which this data was obtained included more than 15,000 survey (questionnaire) respondents representing a broad cross section of Air Force personnel in terms of grades, geographic locations and specialty codes. The survey has six major sections: Background Information, Job Inventory, Supervisory Inventory, Organizational Climate Measures, Work Group Effectiveness Index, and Job Satisfaction Questionnaire. The participants were given a series of statements covering these areas and asked to respond to the following seven point scale:

0 = Not Applicable
1 = Strongly Disagree
2 = Moderately Disagree
3 = Slightly Disagree
4 = Neither Agree nor Disagree
5 = Slightly Agree
6 = Moderately Agree
7 = Strongly Agree

The Variables

Six statements contained in the OAP were selected to measure key dimensions of feedback and performance. The statements "The quality of output of your work group is high" and "The quantity of output of your work group is high" measure perceived quality and quantity of work. The statements "My supervisor lets me know when I'm doing a good job" and "My supervisor lets me know when I'm doing a poor job" measure positive and negative feedback. The statements "My job performance has improved due to feedback received from my supervisor" and "My supervisor frequently gives me feedback on how well I am doing my job" measure quality of feedback and amount of feedback.

The Approach

The study approach was to test six hypotheses (developed from the six statements previously enumerated), stated in the null form, to determine differences between sexes and ethnic groups. There was no attempt to hypothesize about the interaction between sex and ethnicity, but some such interactions were observed and reported for possible use in further studies. In analyzing the data, six two-by-six factorial

design analyses of variance tests for significance were conducted and differences were considered significant at the $p<.05$ level of significance.

Results*

Quality of Work Group Measurement (statement 1, OAP variable 260): On average, males tended to perceive a somewhat higher quality of work group output than did females. Among males, Blacks perceived the highest quality output while Indian/Alaskan natives perceived the lowest quality output. Among females, Whites had the highest quality output perception, and Asian/Pacific Islanders had the lowest. The quality perceptions of white females were higher than those of any other group of either sex. The analysis of variance testing showed significant differences in the variation between the main effects of sex and ethnicity; there was also significant difference in the interaction between sex and ethnicity.

Quantity of Work Group Measurement (statement 2, OAP variable 259): On average, females perceived the quantity of their work group output to be higher than that of males. Among females, Whites had the highest quantity perceptions, and Asian/Pacific Islanders had the lowest. Among males, Blacks had the highest perception of quantity output while Indian/Alaskan Natives had the lowest. White females saw the quantity of their work group output as higher than that of any other group of either sex. Analysis of variance testing showed significant differences in the variation between the main effects of sex and ethnicity. There was no significant difference noted in the interaction between sex and ethnicity.

Positive Feedback Measurement (statement 3, OAP variable 433): On average, females received more positive feedback than males. Among females, Whites received the greatest positive feedback and Indian/Alaskan Natives received the lowest. Among males, Asian/Pacific Islanders perceived the greatest positive feedback, and Indian/Alaskan Natives again perceived the lowest. White females perceived greater positive feedback than any other group of either sex. Analysis of variance testing confirmed significant differences between the main effects of sex and ethnicity; there was no significant interaction between the two effects.

Negative Feedback Measurement (statement 4, OAP variable 434): On average, males perceived more negative feedback than did females. Among males, Whites perceived the greatest negative feedback, and Asian/Pacific Islanders perceived the least. Among females, Hispanics perceived the greatest amount of negative feedback; Asian/Pacific Islanders perceived the least. White males perceived themselves as recipients of more negative feedback than any other group of either sex. Analysis of variance testing again showed significant differences

*Due to space limitation, we were unable to present all demographic and ANOVA tables. However, detailed results and tables are available from the researchers.

in the variation between the main effects of sex and ethnicity. There was no significant interaction between the two effects.

Quality of Feedback Measurement (statement 5, OAP variable 437): This is the only one of the six variables that failed to show any significant differences between the main effects of sex and ethnicity as a result of the analysis of variance testing. As a group, the respondents neither agreed nor disagreed with the statement "My job performance has improved due to feedback received from my supervisor."

Amount of Feedback Measurement (statement 6, OAP variable 442): On average, males perceived greater amounts of total feedback than females. Among males, Asian/Pacific Islanders perceived the greatest amount of feedback, while Whites received the least amount. Among females, Whites received the least amount. Among females, Whites perceived the most total feedback, and Asian/Pacific Islanders perceived the least. Analysis of variance testing confirmed significant differences between the main effects of sex and ethnicity. Using our decision rule, there is also at least a strong inference of interaction between the two main effects with a significant F of .049.

Discussion and Conclusions

Analysis of variance testing confirmed significant differences between the main effects of sex and ethnicity in five of the six hypotheses tested; no significant differences were noted in the quality of feedback measurement (statement 5, OAP variable 437). There also appears to be some significant difference in the interaction between the two main effects for the quality of work group output and the amount of feedback measurements. Generally, the differences between sexes were more pronounced than the differences between specific ethnic groups.

There appears to be a direct correlation between quantity of work and quality of work group performance. White females, for instance, had higher perceptions of both quantity and quality of output than did members of any other group of either sex. Among males, Blacks had the highest quantity and quality perceptions. On the other hand, Indian/Alaskan Natives had the lowest male perception of both quantity and quality of work group output; and Asian/Pacific Islanders had the lowest perception of quantity and quality of output among females. Apparently, the groups which felt they had more to do also felt they produced better quality work; the groups which felt they had less to do felt they produced work of lower quality.

There also appears to be a direct relationship between quality of performance and positive supervisory feedback. White females, the group with the greatest output perceptions, also felt they received the greatest positive feedback. Blacks, who perceived the greatest output quality among males, were the second leading recipients of positive feedback. Indian/Alaskan Natives were the group with the lowest output quality perceptions among males, and they also received less positive feedback than any other male group. The lowest quality perceivers among females, Asian/Pacific Islanders, received less positive feedback than all but one other group (Indian/Alaskan Natives).

BIBLIOGRAPHY

Glover, J.G. Fundamentals of professional management. New York: Simmons Boardman Publishing Corp, 1958.

Hodgetts, R.M. Management: Process and practice. Philadelphia: W.B. Saunders Company, 1975.

Kirby, P. Productivity increase through feedback systems. Personnel Journal, 1977, 56, 512-515.

Pritchard, R.D., Montagno, R.V. & Moore, J.R. Enhancing productivity through feedback and job design. (AFHRL-TR-77-56). Brooks AFB, TX: Air Force Human Resources Laboratory, 1978.

USAF, Air University LMDC management consultation information system report #1: Executive leadership and management summary. Maxwell AFB, Alabama, 1978.

Motivation in the Aircraft Armament Systems Career Field

Jerry R. Pfeifer, Major, USAF
Air Command and Staff College
Lawrence O. Short, Captain, USAF
Air Force Leadership and Management Development Center
Maxwell Air Force Base, Alabama

Abstract

Declining reenlistment rates and prior studies indicating low job interest highlighted motivational problems in the Air Force Armament Career field. Analysis of perception of job factors revealed significant lower ratings by armament personnel, compared to other maintenance personnel. A possible lack of interaction between individuals, supervision, and the organization existed, apparently at least in part due to poor communications practices. Improved management practic were recommended.

The motivation of Aircraft Armament Systems maintenance personnel directly effect the combat readiness of an operational Air Force unit. It ensures proper weapon performance, and airborne gun and weapons release system operation.

Previous studies of the armament career field identified perceptions of low job interest and underutilization of talents and training (Gould, 1972; Barucky & Burns, 1976). Reenlistment rates for first-term armament personnel dropped from 38.6 percent in Fiscal Year (FY) 76 to 24.1 percent in FY 79, compared to overall Air Force rates of 37.3 percent and 38.0 percent respectively.

Recruiting efforts were limited by the non-transferrability of the skill to a civilian occupation and lack of a monetary bonus program for the career field (Keasling, 1979). With an average initial training cost of $4,650 per student (Barucky & Burns, 1976, p.35), maximum retention through management initiatives is imperative.

Since motivation largely depends on the perceptions a worker has toward his job or organization (Steers & Porter, 1975), we formulated the hypothesis that job and organizational perception of enlisted personnel in the Aircraft Armament Systems career field were significantly different from other enlisted maintenance personnel. The hypothesis was tested against the following job factors:

800	Skill Variety	814	Work Repetition
801	Task Identity	816	Desired Repetitive/Easy Tasks
802	Task Significance	817	Advancement/Recognition
804	Job Feedback	818	Management/Supervision
805	Work Irritants	819	Supervisory Communications Climate
806	Need for Enrichment	820	Organizational Communications Climate
810	Job Performance Goals	821	Perceived Productivity
811	Pride	822	Job Related Satisfaction
812	Task Characteristics	823	Job Related Training
813	Task Autonomy	824	General Organizational Climate

Background

The armament career field consists of approximately 11,500 people, with 94 percent assigned to four Air Force Major Commands (Nowlin, 1980). The majority of duty assignments are at wing or squadron level. Duty assignments fall in one of three functional areas: (1) Weapons Loading, responsible for loading, unloading, arming and disarming of committed munitions, (2) Weapons Release, responsible for testing, inspection, scheduled and unscheduled maintenance and repair of aircraft weapons release systems, and (3) Gun Services, responsible for maintenance, inspection, removal and installation of airborne gun systems. (Department of the Air Force, 1973). The entire career field averages 28 percent of available manhours loading and unloading munitions, 29 percent on supervisory and administrative duties, and the other 43 percent on inspection, repair, or training (Barucky & Burns, 1976, p. 14).

Motivational factors were considered from the systems perspective, where behavior is controlled by the individual, the environment, and feedback mechanisms that redirect goals. No single factor controls motivation (Steers & Porter, 1975, p. 6). The instrument used to measure individual perceptions of the organization, the job, and management is a 109 question survey called the Organizational Assessment Package (OAP). The OAP was designed by the Air Force Leadership and Management Development Center (LMDC), Maxwell AFB, AL, and the Air Force Human Resources Laboratory, Brooks AFB, TX, to provide a means of identifying existing strengths and weaknesses within organizational work groups and aggregated work groups (Hendrix and Halverson, 1979a, 1979b). An important part of the LMDC mission is to provide consultant services to Air Force Commanders on request, and to provide leadership and management training to personnel in their work environments. Following consulting visits, OAP responses are added to the OAP data base at Maxwell AFB, AL for research use.

Method

The data base used for this study consisted of the responses of approximately 2,300 enlisted maintenance personnel, of which approximately 190 were armament personnel. Data were grouped for armament and other maintenance personnel, and means and standard deviations were computed for each factor for each group. A student's "t" test was applied to each set of responses to identify significant differences.

Results

Results are shown in the following Table. Significant differences existed between perceptions of the two groups for all but three of the 20 factors at the 95% level of confidence or greater.

TABLE 1

OAP DATA ANALYSIS RESULTS

FACTOR	GROUP						t-VALUE
	ARMAMENT			OTHER MAINTENANCE			
	N	X	SD	N	X	SD	
800	189	3.8175	1.433	2,154	4.4773	1.455	-5.98***
801	190	4.9579	1.271	2,144	5.1474	1.224	-2.04*
802	190	5.3974	1.492	2,164	5.5039	1.373	-1.02
804	190	4.1395	1.467	2,160	4.7549	1.282	-5.60***
805	184	4.6466	1.125	2,105	4.7230	1.101	-0.90
806	188	4.8511	1.555	2,135	5.3745	1.284	-4.48***
810	186	4.4247	1.035	2,134	4.7399	0.978	-4.20***
811	187	3.7727	1.771	2,155	4.6093	1.715	-6.38***
812	187	4.5745	1.009	2,120	4.9744	0.976	-5.35***
813	184	2.5856	1.362	2,127	3.6695	1.414	-10.00***
814	190	5.3553	1.364	2,148	4.9758	1.399	3.59***
816	190	3.3684	1.482	2,118	3.1249	1.346	2.19*
817	188	3.7362	1.286	2,102	4.1441	1.189	-4.48***
818	182	4.3573	1.609	2,119	4.9258	1.510	-4.85***
819	187	4.0757	1.573	2,113	4.4535	1.548	-3.19**
820	185	3.9188	1.375	2,094	4.3560	1.300	-4.36***
821	188	5.3713	1.296	2,115	5.5553	1.181	-1.88
822	166	3.8704	1.305	1,931	4.6706	1.323	-7.49***
823	180	4.1028	1.602	2,065	4.4344	1.569	-2.71**
824	183	3.8645	1.402	2,089	4.2705	1.418	-3.72***

* p <.05
** p <.01
*** p <.001

Discussion

Armament personnel hold significantly different, and generally lower, perceptions of their jobs, supervisors, and organizations than other maintenance personnel.

Job Perceptions. Armament personnel perceive less task autonomy (813), skill variety (800), and task identity (801) than other maintenance personnel, indicating a feeling that jobs are structured and that priorities are determined by others. The indicated higher desire for repetitive tasks (816) has apparently been met, as indicated by a perception of high work repetition (814). Armament personnel view their jobs as significant (802) and perceive their productivity essentially the same as other maintenance personnel. They perceive less need for enrichment (806), get less non-verbal feedback (812), and feel less pride (811) than other maintenance personnel.

Perception of Supervisors. Armament personnel perceived significantly less supervisory ability and interaction (818) and communication (819), perhaps reflecting a more authoritarian approach to leadership. Low perceptions of job performance goals (810), verbal job feedback (804) and job related training (823) indicated a weakness in worker-supervisor relationships. The lower perception of job satisfaction (822) and pride (811) by armament personnel can possibly be attributed to the apparent low level of supervisory interaction.

Perceptions of Organization. No significant difference existed between groups regarding work irritants (805), perhaps indicating that similar physical work conditions exist. Armament personnel perceived less organizational involvement in their advancement, and less chance of having good performance rewarded (817).The apparent result was a lower regard for armament organizations (824). Armament personnel also perceived a lower level of communication within the organization (820), which could also have been reflected in several of the other factors.

Conclusions

Armament personnel perceived a lower need for enrichment and a higher desire for repetitive tasks. Communication channels appeared to be obstructed at most levels, resulting in low feedback, lack of clear job performance and organizational goals, and possible individual frustration at not being able to identify fully with the unit mission. This seems especially important in light of the finding that workers today find little value in jobs that are menial and nonchallenging, and seek greater interaction with their environment (Lawler, 1973).

BIBLIOGRAPHY

Barucky, J.M. & Burns, P.C. Weapons mechanic career ladder: AFSC 46230, 46250, 46270, and 46290. (AFPT 90-46X-052) Lackland AFB, TX: Air Force Occupational Measurement Center, November 1976.

Department of the Air Force, Air Force manpower standards, mission equipment maintenance and operations (Air Force Manual 26-3, Vol. III). Washington, D.C. 1973.

Gould, R.B. Reported job interest and perceived utilization of talents and training by airmen in 97 career ladders (AFHRL Rep. TR-72-2). Brooks AFB, TX: Air Force Human Resources Laboratory, January 1972.

Hendrix, W.H. & Halverson, V.B. Organizational survey assessment package for Air Force organizations (AFHRL-TR-78-93). Brooks AFB, TX: Air Force Human Resources Laboratory, February 1979. (a)

Hendrix, W.H. & Halverson, V.B. Situational factor identification in Air Force organizations (AFHRL-TR-79-10). Brooks AFB, TX: Air Force Human Resources Laboratory, May 1979. (b)

Hughes, C.L. Goal setting: key to individual and organizational effectiveness. New York: American Management Association, 1965.

Keasling, Master Sergeant D., Branch Chief, Retention Training Programs, Assistant for Enlisted Retention Air Force Manpower and Personnel Center, Randolph AFB, TX, Telephone Interview, 28 November 1979.

Lawler, E.E. III. Motivation in work organizations. Belmont, CA: Wadsworth Publishing Co., Inc., 1973.

Nowlin, Senior Master Sergeant C., NCOIC Avionics and Weapons System: Assignment Section, Air Force Manpower and Personnel Center, Randolph AFB, TX, Telephone Interview, 7 January 1980.

Steers, R.M. & Porter, L.W. Motivation and work behavior. New York: McGraw-Hill, Inc., 1975.

Equal Opportunity and Treatment: Perceptions of Air Force Military Men and Women

Charles R. Spicher, Major, USAF
Air Command and Staff College

Kenneth L. Hamilton, Lt Col, USAF
Air Force Leadership and Management Development Center
Maxwell Air Force Base, Alabama

ABSTRACT

While reported cases of sexual discrimination within the Air Force are on the decline, this paper examines a hidden potential problem: differing perceptions of Air Force military men and women concerning opportunities and treatment within the Air Force. Based on analysis of variance (ANOVA) of the Leadership and Management Development Center's Organizational Assessment Package (OAP) data base, fourteen of twenty-three job factors were found to display statistically significant differences between sex groups. These factors pertain to job enrichment, the nature of the duty tasks, advancement/recognition, and organizational climate. The disparities in the perceptions of Air Force military men and women suggest possible areas for focus of Equal Opportunity and Treatment/Human Relations Education program emphasis and Affirmative Actions Plan monitoring.

The perceptions of Air Force military men and women concerning their opportunities for a rewarding career and treatment commensurate with their peers are vital to an efficient and functional Air Force. The differing perceptions held by the members do not have to be based on an actual situation, only a perceived belief, to be detrimental to the organizational health of the unit. For a perception, fact or fiction, is truth to the perceiver.

Differences in the perceptions of military men and women are not new to the Air Force. A 1976 Air Force Department of Social Actions Training Survey of 26,000 military personnel revealed: a) women have more positive attitudes than men about the capabilities of women to perform on an equal level with men; b) women perceive more sexist behavior than men; and c) discrimination towards women on the job appears significantly greater than that directed toward other minority groups. A second 1976 Air Force survey of military personnel revealed: a) forty percent of the enlisted force and twenty-five percent of the officer force believed men and women were not given the same responsibilities on the job, and b) two-thirds of the enlisted force and one-half of the officer force believed that men and women were not treated equally. (Is Being a Woman Helpful or Harmful, 1977) A more recent study documented 80 separate observations on 43 different Air Force bases where women perceived the Air Force as a man's Air Force, lacking equal opportunity and treatment, with jobs tailored to the needs and

desires of men, and discriminatory practices in job assignments. (LaNier, 1979)

Several reasons for these differing perceptions may exist, but one of note is sex role stereotyping. This factor was highlighted in a University City Science Center study, conducted under contract to the Navy, to investigate in a sample similar to the enlisted force of the Navy the differences between men and women's perceptions and attitudes in the work environment. (Good, Kirkland, & Grissom, 1979) To combat this "learned from childhood" sex role stereotyping, education has been offered as one solution. (MacCoby & Jackling, 1974, pp. 73-83 and Wesley & Wesley, 1977, p. 1)

In 1971, the Air Force instituted the race relations education program to create "awareness" of discriminatory practices. The major problem was that the program was directed only at racial issues since sex has never caused any riots as has race issues. (Sanchez, 1978, p. 28) The history of the Air Force Race Relations Program, since renamed Human Relations Education (HRE), shows a decline in emphasis from a mandatory 18 hour course in 1971 to a short briefing each time a military member arrives at a new duty station or attends Professional Military Education courses. The problem remains though that education to reduce sex discrimination, at its highest point, was only two hours in the third generation HRE course, Discipline and Human Relations, and directed only at supervisors and managers, not the Air Force as a whole.

The lack of education in sex-discrimination is further complicated by the increase in the number of women in the Air Force from one percent (9,858) in FY 67 to 8.3 percent (41,084) in FY 78. The total number of women is expected to reach about 98,000 by FY 85 or about 17 percent of the total force. (USAF Statistical Digest, 1979) While it is true the total number of reported sex discrimination cases declined from 407 in 1976 to 286 in 1978 (Love, 1979) during a period of increasing numbers of women in the Air Force, the area of perceptions was only starting to be studied.

Method

The OAP data base utilized contained the responses of over 16,000 military men and women. The men accounted for 90.8 percent of the sample and the women 9.2 percent. These percentages compared favorably to the current total Air Force ratio of 90.7 percent men and 9.3 percent women. (Lester, 1980) Analysis was performed using a sample analysis of variance on the following factors:

800	Skill Variety	814	Work Repetition
801	Task Identity	816	Desired Repetitive/Easy Task
802	Task Significance	817	Advancement/Recognition
804	Job Feedback	818	Management/Supervision
805	Work Irritants	819	Supervisory Communication Climate
806	Need for Enrichment	820	Organizational Communication Climate
807	Job Motivation	821	Perceived Productivity
808	Job Inventory	822	Job Related Satisfaction
809	Job Motivation Index (additive)	823	Job Related Training
810	Job Performance Goals	824	General Organizational Climate
811	Pride		
812	Task Characteristics		
813	Task Autonomy		

Results

Results of the analysis of variance and computed means and standard deviations are shown in the Table 1. Significant differences existed between the perception of the two groups in 14 of the 23 job factors at the 95% or greater confidence level.

TABLE 1: OAP DATA ANALYSIS RESULTS

	MALE			FEMALE			F VALUE
FACTOR	N	X	SD	N	X	SD	F
800	11495	4.50	1.55	1155	3.98	1.50	19.47***
801	11495	4.91	1.36	1155	4.77	1.30	0.44
802	11495	5.47	1.50	1155	5.45	1.41	24.27***
804	11495	4.64	1.36	1155	4.54	1.35	2.05
805	10818	4.53	1.14	1085	4.67	1.17	27.94***
806	10818	5.47	1.30	1085	5.54	1.21	46.93***
807	10818	98.90	64.59	1085	92.57	60.65	17.29***
808	10818	64.26	13.04	1085	63.13	12.33	19.56***
809	10735	13.42	3.12	1073	13.13	3.01	19.66***
810	10735	4.62	1.06	1073	4.53	0.99	1.65
811	10735	4.63	1.78	1073	4.45	1.78	3.62
812	10735	4.89	1.11	1073	4.69	1.05	1.19
813	10935	3.90	1.48	1090	3.85	1.41	32.59***
814	10935	4.86	1.45	1090	5.10	1.43	5.69*
816	10935	3.07	1.43	1090	3.12	1.37	4.76*
817	10935	4.21	1.25	1090	4.06	1.19	6.04*
818	10448	4.97	1.54	1051	4.88	1.61	0.19
819	10448	4.55	1.58	1051	4.51	1.61	0.73
820	10448	4.39	1.39	1051	4.36	1.34	4.13*
821	10448	5.47	1.28	1051	5.31	1.28	0.04
822	9173	4.78	1.35	899	4.80	1.25	24.15***
823	9123	4.39	1.61	899	4.30	1.61	0.18
824	9173	4.44	1.48	899	4.35	1.45	5.75*

*$p < .05$ **$p < .01$ ***$p < .001$

Discussion

Air Force military women do have statistically significant different perceptions, generally lower, of 14 of the OAP job factors. There are five factors where the women responded at a higher level than did the men; however, two of these could be considered negative factors. Women believe they are performing more repetitive types of tasks (814), but they indicate a desire for easy/repetitive types of tasks (816). A word of caution--the level of desire for the repetitive tasks by women is at a very low level and should not be considered a compensating effect for the repetitive nature of their tasks (814). The women responded at a higher level than the men for job related satisfaction (823). Women perceive fewer work irritants (805), and indicate a higer potential for enrichment in their jobs (806) than do the men. With these conditions in mind, we can say that the men generally responded at a higher level or had a better perception (negative factors) of the OAP factors in 11 of the 14 significantly different areas.

In addition to the factors mentioned above, men generally experience greater variety in the skills required and complexity of their duties (800). Men believe their jobs are important to others and the mission (802) more so than do the women. Men have a stronger motivation to the job (307) and experience more autonomy (813) than do their female counterparts. The communication climate within the organization (820) and the general organization takes care of men better than the women (824). A primary area of concern is advancement/recognition (817). Again, the men responded higher, raising the possibility that the women are not aware of their opportunities to advance, are not being groomed or trained for advancement, and/or do not believe their efforts/output will be recognized as readily as their male counterparts.

However, there were nine factors where no significant differences occurred. In this area, the men and women both perceived the factors in a similar manner. Both groups have an identity with their job/tasks (801), receive sufficient feedback from the job itself to let them know how they are doing (804), have similar job performance goals and assistance in setting realistic and attainable goals (810), have a sense of pride in their work (811), see similar characteristics in their jobs (812), believe their supervisors take care of them and are receptive to their ideas (818), can talk to their supervisors (819), their work group's output is high and resource conscious (821), and have received the necessary training, both technical and on the job, to accomplish assigned duties (823).

Conclusions

Military women do not have the same perceptions as military men in over 60 percent of the job factors evaluated in the OAP. For the job factors where statistically significant differences did occur, the men responded at a higher level or had a better perception than did the women in 11 of the 14 significantly differing areas.

BIBLIOGRAPHY

Air Force Social Actions Training Department. An evaluation of the Air Force race relations training program. Randolph AFB TX: USAF Social Actions Program Analysis Branch, April, 1976.

Good, J., Kirkland, F. R., & Grissom, G. R. Working relationships between men and women: Effects of sex and hierarchical position on perceptions of self and others in a work setting (Technical Report Number 6). Philadelphia, PA: University City Science Center, 1979.

Is being a woman harmful or helpful? Air Force Times, June 10, 1977, p. 26.

LaMier, J. V., Leadership and management problems. Unpublished Research Report, Air War College, Maxwell AFB AL, 1978.

Lester, M., When the boss is a woman. The Times Magazine Supplement to Army Times/Navy Times/Air Force Times. January 14, 1980, pp. 4-10.

Love, Captain J. E., Social Actions Office (AFMPC/MPCXS), Air Force Manpower and Personnel Center, Randolph AFB TX, Telephone Interview, November, 1979.

MacCoby, E., & Jacklin, C., The psychology of sex differences. Stanford CA: Stanford University Press, 1974.

Sanchez, J. R., More women in the Air Force: Implications for Air Force supervisors. Unpublished Research Report, Air War College, Maxwell AFB AL, 1978.

USAF statistical digest, FY78 (s) (RCS: HAF-ACM(a) 7203). Washington DC; Comptroller of the Air Force, May 31, 1979.

Wesley, F., & Wesley, C., Sex role psychology. New York: Human Sciences Press, 1977.

An Assessment and Comparison of Air Force Job Related Satisfaction by Air Force Personnel Categories

Glenn E. Boren, Major, USAF
Air Command and Staff College
Lawrence O. Short, Capt, USAF
Air Force Leadership and Management Development Center
Maxwell Air Force Base, Alabama 36112

Abstract

A portion of the Air Force's Organizational Assessment Package (OAP) data base was used to assess and compare job related satisfaction for four Air Force Personnel Categories. Factors measuring variables of five job core characteristics (skill variety, task identity, task significance, task autonomy, task feedback) were extracted from the OAP. Analysis of variance and multiple comparisons were performed by factor to identify significant differences between personnel categories. Motivating Potential Scores (MPS) were calculated for each personnel category. Significant differences by job factor and a comparison of MPSs for each personnel category are presented. Implications for future AF job enrichment efforts and for recruiting and retention programs are discussed.

In order to accomplish its assigned mission in providing for the national security of the United States, the Air Force must recruit, train, effectively use, and retain a sufficient workforce of officer, enlisted, and civilian personnel. Many high-level officials in the defense community (Allen, 1979; Brown, 1979; Pirie, 1980) have identified the problems of recruitment and retention as the biggest challenge facing the armed services in the 1980's. Complicating the problem is the fact that the abolishment of the draft in 1973 has forced the services to compete directly with the private sector in trying to keep the All-Volunteer Force fully manned.

This study suggests that "job satisfaction" within the Air Force work force may offer at least a partial solution to the Air Force's recruitment and retention problem of the 1980's. It is a widely accepted premise (Hardy, 1978; Morgan, 1973) that the military cannot compete directly with the private sector solely on the basis of extrinsic job incentives such as pay and benefits. Therefore, the intrinsic factors providing motivation and satisfaction may offer the only solution. Some behavioral scientists (Herzberg, Mausner, and Snyderman, 1959; Ford, 1969; Herzberg, 1977) have shown dedicated, satisfied workers to be generally more efficient and productive. Others, such as Kraut (1975) have shown that workers tend to migrate toward jobs which they expect offer greater personal satisfaction and remain in those jobs which deliver the satisfaction they expect. Military researchers (Guinn and Truax, 1973; Patterson, 1977) have shown these same intrinsic factors at work in those seeking and holding jobs in the Air Force. However, before suggesting that Air Force recruitment and retention programs be restructured around job satisfaction incentives, an

PRECEDING PAGE BLANK-NOT FILMED

assessment of job satisfaction within the Air Force's present workforce was necessary. Such an assessment is the purpose of this study.

Method

In 1975, Hackman, Oldham, Janson, and Purdy proposed that intrinsic job motivation and satisfaction were a function of five measurable job core characteristics. skill variety, task identity, task significance, feedback, and autonomy. The measureable values of these characteristics could also be used to calculate a Motivating Potential Score (MPS) using the following formula:

$$MPS = \frac{\text{Skill Variety} + \text{Task Identity} + \text{Task Significance}}{3} \times \text{Feedback} \times \text{Autonomy}$$

The MPS provides a quantitative measure of the amount of motivation and satisfaction being provided a job's incumbent and a measure of the potential for enrichment.

The Air Force's Organizational Assessment Package (OAP) is a survey used by the Leadership and Management Development Center (LMDC) and contains measures (on a 7 point Likert-type scale) of factors identical to the Hackman-Oldham core characteristics. OAP results are computerized and can be accessed by numerous factors and partitioned in various ways.

Subjects for the study were approximately 16,200 Air Force Officers (2400), enlisted airmen (10,600), General Schedule civilians (2500), and Wage Board civilians (700). The data base compared generally to the 1979 Air Force distribution of these personnel categories.

The OAP data base was partitioned by the four personnel categories for the five job core characteristics. Survey response means and standard deviations were obtained for each personnel category for each job factor (core characteristic). A one-way analysis of variance followed by a Newman-Keuls multiple comparison test was performed for each factor.

Finally, using the mean scores for each personnel category for each job factor, a MPS was calculated for each personnel category using the Hackman-Oldham formula. The MPS provided the basis for comparing overall job satisfaction between personnel categories.

Results

TABLE 1 - Skill Variety Factor

Group	N	Mean	Standard Deviations
Officers	2460	5.09	1.42
Enlisted	10,634	4.30	1.55
GS Civ	2559	4.92	1.41
WB Civ	734	5.11	1.35

Table 1 results showed significant differences existed between some categories for this factor ($F=288$; $df=3,16383$; $p< .001$). Results of the Newman-Keuls ($p< .05$) showed that the score for enlisted personnel was significantly lower than for any other groups for the Skill Variety Factor. The scores for officers and Wage Board civilians were significantly higher than for other groups, though not significantly different from each other. The scores of the General Schedule civilians were significantly different from those of other groups and occupied the mid-range for this factor.

TABLE 2 - Task Identity Factor

Group	N	Mean	Standard Deviations
Officers	2444	5.05	1.27
Enlisted	10,573	4.87	1.38
GS Civ	2554	5.32	1.20
WB Civ	733	5.30	1.20

Table 2 results showed that significant differences existed ($F=97$; $df=3,16300$; $p< .001$). The Newman-Keuls test ($p< .05$) showed the task identity score for enlisted personnel to be significantly lower than the other groups. The scores for General Schedule and Wage Board civilians were significantly higher, although not significantly different from each other. The task identity score for officers was significantly different from all other groups with its value falling between the high and low.

TABLE 3 - Task Significance Factor

Group	N	Mean	Standard Deviations
Officers	2470	5.52	1.45
Enlisted	10,667	5.44	1.50
GS Civ	2563	5.65	1.30
WB Civ	727	5.72	1.22

Table 3 results showed significant differences for this factor ($F=21$; $df=3,16423$; $p< .001$). The Newman-Keuls ($p< .05$) test showed the task significance scores of enlisted personnel to be signficantly lower than the scores for any other group. The scores for General Schedule and Wage Board Civilians, though not significantly different from each other, were significantly higher than all other groups. The scores for officers were significantly different from the other groups and were the mid-range scores for the Task Significance Factor.

TABLE 4 - Job Feedback Factor

Group	N	Mean	Standard Deviations
Officers	2465	4.71	1.26
Enlisted	10,661	4.61	1.38
GS Civ	2564	5.03	1.26
WB Civ	729	5.11	1.27

Table 4 results showed significant differences ($F=91$; $df=3, 16415$; $p< .001$). The Newman-Keuls test ($p< .05$) showed the enlisted to again be significantly lower. GS and WB civilians were again significantly higher, though not significantly different from each other. Officers were significantly different and in the mid-range of scores for Feedback.

TABLE 5 -Task Autonomy Factor

Group	N	Mean	Standard Deviations
Officers	2438	4.42	1.40
Enlisted	10,504	3.78	1.46
GS Civ	2541	4.59	1.37
WB Civ	727	4.47	1.32

Table 5 results again showed significant differences to exist ($F=321$; $df=3, 16206$; $p< .001$). The Newman-Keuls test ($p< .05$) showed the autonomy score for enlisted personnel to be significantly lower than for other groups. The scores for GS civilians were significantly higher than the other groups. There was no significant difference in the scores for officers and WB civilians for the Autonomy Factor, even though they did significantly differ from other groups and fell in a mid-range.

The MPSs calculated for each personnel category using the Hackman-Oldham formula yield the following result: Officers = 108.67, Enlisted = 84.86, GS Civilians = 122.36, WB Civilians = 122.89.

DISCUSSION AND CONCLUSIONS

Based on the fact that measurable significant differences were found for each core characteristic related to job satisfaction, we can conclude that there are significant differences in the degree of job satisfaction being provided by the Air Force to its various personnel categories. Based upon the calculated MPSs, the Wage Board Civilians are experiencing the highest level of job satisfaction, followed closely by the General Schedule Civilians. Air Force officer job satisfaction falls at about the level of the average for the total work force. The job satisfaction being experienced by Air Force enlisted personnel is the lowest for any of the four personnel categories.

The motivation behind the study was to determine if Air Force jobs are satisfying to its work force and enough so to be used as motivators in recruitment and retention programs. For officers and civilian personnel the results appear to show a good degree of job satisfaction and enough so as to act as incentives in recruitment and retention. For enlisted personnel, the results are not as conclusive and there would appear to be room for improvement in the areas of "skill variety" and "autonomy". This would appear to coincide with findings of an Air Force Magazine survey (Gates, 1980) of those who once separated from the Air Force and have since returned to duty. "Job Satisfaction" and "importance of the mission" were cited as top reasons for returning.

BIBLIOGRAPHY

Allen, L. Role of the commander and supervisor in retention. TIG Brief, 1979, 31, 13.

Brown, H. Department of Defense annual report: Fiscal year 1980. Washington, DC: U.S. Government Printing Office, 1979.

Ford, R. Motivation through the work itself. American Management Association, 1969.

Gates, E. The other side of the retention: Those who return. Air Force Magazine, 1980, 63, 62-65.

Guinn, N. and Truax, S. Comparison of volunteer attitudes and career motivation among officer and airman personnel. (AFHRL-TR-73-28). Brooks AFB, TX: Air Force Human Resources Laboratory, October 1973.

Hackman, J.R., Oldham, G.R., Janson, R., and Purdy, K. A new strategy for job enrichment. California Management Review, 1975, 17, 57-71.

Hardy, J.N. Sociological change: Impact on the all-volunteer force. Maxwell AFB, AL: Air War College, April 1978.

Herzberg, F., Mausner, B., Snyderman, B.B. The motivation to work. New York: Wiley and Sons, 1959.

Herzberg, F.I. Orthodox job enrichment: A common sense approach to people at work. Defense Management Journal, 1977 13, 21-27.

Kraut, A.I. Predicting turnover of employees from measured job attitudes. Organizational Behavior and Human Performance, 1975, 13, 233-243.

Morgan, H.S. Job satisfaction: Key to motivation of the Marine Corps' "few good men." Maxwell AFB, AL: Air Command and Staff College, May 1973.

Patterson, J.W. An analysis of career intent and job satisfaction of first term Air Force personnel. Wright-Patterson AFB, OH: Air Force Institute of Technology, School of Engineering, December 1977.

Pirie, R.B. The continuing military manpower crunch: Defense 80 Washington, DC: U.S. Government Printing Office, January 1980.

CONCEPTUAL MODEL OF A MERIT PAY SYSTEM FOR AIR FORCE CIVILIAN EMPLOYEES

John A. Querrieri, Douglas K. Cowan, Bart M. Vitola
Air Force Human Resources Laboratory
Brooks AFB, TX 78235

Abstract

The Civil Service Reform Act of 1978 (PL 95-454) has levied the requirement for government agencies to distribute merit pay to GS 13, 14, and 15 supervisors and management officials. This paper will discuss the conceptual model developed by AFHRL to evaluate the performance of individuals against pre-determined standards, and the method for converting these individualized ratings to more useful normative ratings. It will also discuss possible methodologies for the distribution of salary increases.

The passage of Public Law 95-454 has generated a substantial amount of concern and interest among all federal employees. This paper will deal with that aspect of PL 95-454 which pertains to the merit pay system.

Title 2 of the CSRA 5USC 4302A states that "each agency shall develop one or more performance appraisal systems which: 1) provide for periodic appraisals of job performance of employees; 2) encourage employee participation in establishing performance standards; and 3) use the results of performance appraisals as a basis for training, rewarding, reassigning, promotion, reducing in grade, retaining, and removing employees". In addition, Title 5 of the CSRA (5USC 5401) states in part that the purpose of the merit pay section is to provide for a system which shall "within available funds, recognize and reward quality performance by varying merit pay adjustments; use performance appraisals as the basis for determining merit pay adjustments...apply to any supervisor or management official...who is in a position which is in GS-13, 14, or 15 of the General Schedule". The individuals who meet the merit pay definition will be under a separate pay system and be categorized as GMs (General Managers). It is for this group of people that the General Manager Appraisal System (GMAS) is being developed.

Historically, less than one percent of all federal civilian employees who are eligible for within-grade pay increases have had their increases withheld. At the same time, less than two percent per year of all federal emoloyees received salary increases (Quality Step Increases) for outstanding performance. Also, the current system requires management to provide a preponderance of evidence to support personnel actions such as removal. Under the CSRA procedures, in the case of an action based on unacceptable performance, management is required to produce substantial evidence to support their case. This particular portion of the law and those referenced above should dispel the myth that it is almost impossible to remove proven non-producers or effectively reward outstanding performances in the federal services.

Individuals identified as GMs are considered to be the feeder pool

PRECEDING PAGE BLANK-NOT FILMED

for the Senior Executive Service (SES). Therefore, GMAS was conceptualized to incorporate some of the salient aspects of the Senior Executive Appraisal System in addition to those features unique to a merit pay system.

The Merit Pay System described in this paper is a conceptual model subject to revisions based on field tests and logistic necessities. The GMAS has two major components: The Appraisal Process and the Merit Pay Review Board.

The Appraisal

The appraisal component has three major segments: 1) the workplan, 2) workplan/performance reviews, and 3) supervisory ratings. The workplan is designed to provide the GM and his/her supervisor a medium for delineating the job elements and performance standards specific to the individual GM's position. The CSRA specifically states that performance standards and critical elements must be determined at the beginning of the rating period, and the employee is encouraged to participate in their determination.

In designing the workplan, a literature review was performed to identify a set of core functions that could be used by all GMs regardless of specific type of work performed. Twenty-five candidate functions were identified. The functions were listed and sent to 500 randomly selected GS 13 through 15 Air Force supervisors with instructions to rank 10 of the 25 that were most pertinent to the employee's position. Based on the results of the rankings and the combination of specific subareas into broader areas of responsibility, the workplan was divided into six functional areas. These areas are: 1) administration; 2) communication; 3) evaluation; 4) planning; 5) reviewing; and 6) technical applications.

Although functions were now established that would standardize the rating form across job series, it was determined that the functions alone would not suffice to allow a supervisor to provide an objective performance appraisal on an employee. Therefore, under each function, the provision was made for the supervisor and employee to develop specific job elements that could be tailored to individual performance, while maintaining the structure provided by standardized functions.

In an effort to facilitate objective performance appraisals to the maximum extent possible, a method was provided to allow the supervisor and employee to establish performance standards for a fully successful employee against which the employee could be compared. This was accomplished by providing space for the standards to be entered on the form. However, the method of determining the standards did pose a problem. Since it was important to provide standards for a function or for elements subsumed under that function, it became necessary to provide for multi-faceted standards.

It is possible that not all functions specified on the rating form are required to adequately delineate an employee's job. Therefore, a method was devised to weight the functions. The weighting is achieved by having the supervisor and employee distribute percentages among the functions. Thus, for however many functions were selected, the total percentage (100) would account for the total job.

Another item mandated by Public Law 95-454 was the requirement for

the supervisor to identify at least one critical element of an employee's job. Although this is a supervisor's responsibility, joint determination by employee and supervisor is encouraged. However, since several elements may appear under one function and only one of the several elements may be critical, it was decided that if any element was declared critical, this would cause the function to be critical. Therefore, at least one function must be designated as critical.

The flexibility inherent in the workplan combined with the structure furnished by the standardized functions provide the GMAS the foundation for designing a methodology for converting an individualized rating (whereby an employee is rated against a set of predetermined, individualized standards) to a normative ranking (whereby an employee is compared with his/her peers).

Because of the types of jobs being considered, it is vital that the appraisal system be dynamic and allow for change and feedback during the rating period. The periodic workplan/performance review fulfills this need. During the review, the GM and supervisor meet and identify any changes required in the workplan and discuss the employee's performance for that period. The supervisor would indicate how the GM is performing relative to each standard, and the employee would have the opportunity to bring job-related problems to the supervisor's attention. These discussions would ensure that the employee would not be surprised when the final rating is given, and they would provide the supervisor with documentation to support recommendations for merit pay raises or other personnel actions.

At the end of the appraisal period, the supervisors would then rate the GM employees on their performance on each function and their overall performance. This difficult and usually subjective rating task becomes significantly easier and more objective because of the workplan. The workplan provides the methodology for specifying job elements and establishing performance standards within the structure of specified functional areas. The workplan/performance reviews allow both supervisor and employee to track the performance of the employee; therefore, the rating official would be able to quickly complete the employee's performance rating at the end of the rating period.

Since the standards are based on the expected performance of the conscientious, productive employee, a rating scale that provides differentiation between various performance levels was required. The scale developed has five ratings: Unacceptable, Marginally Acceptable, Fully Successful, Exceptional, and Superior.

After rating the individual functions, the supervisor will rate the GM employee's overall performance using the same scale. The overall performance rating would be used to qualify the employee for merit pay or initiate other personnel actions. The overall rating was designed to allow the supervisor to take into account the individual function ratings and to consider any other factors that may have influenced the employee's overall performance. Therefore, the rating need not necessarily be a weighted average of the function ratings. However, if any critical functions were performed in an unacceptable manner, then the highest allowable overall rating would be "Marginally Acceptable". Operationally

defined, a rating of Marginally Acceptable means that the employee is retainable but may not be recommended for a merit pay adjustment for this rating period. Further, in the event of an Unacceptable or Superior rating, the supervisor would be required to furnish written, objective justification, especially if the overall rating varied considerably from what would be expected from viewing the function ratings.

The last, and perhaps most important, requirement is the decision by the supervisor to recommend or not recommend the employee for a merit pay increase, and the concurrence or non-concurrence of the reviewing official (the rating official's supervisor). Once the rating is completed and reviewed, it would be sent to the Merit Pay Review Board (MPRB).

The Merit Pay Review Board

The concept of the MPRB is perhaps the most innovative and sophisticated component of the Merit Pay System. The purpose of the MPRB is to: 1) normalize the individualized ratings by rating the workplans on a set of predefined criteria; 2) minimize personal bias; 3) increase user acceptance; 4) provide appropriate minority and organizational representation; 5) provide the variance necessary to accurately determine an order of merit based on performance; and 6) ensure management participation.

Past experience indicates that the selection board process has provided satisfactory results at both the senior enlisted and officer levels in the United States Air Force. Therefore, the selection board process was considered a viable method to determine merit pay increases for DAF civilian employees. There will be a selection board for each grade level to ensure that grade will not influence the decisions of the board. The Merit Pay Review Board shall: 1) be convened at a predetermined organizational level; 2) be composed of proportional representation for minorities, organizations, and job series; 3) consist of an appropriate number of three-member panels and a board president of a higher grade than those being considered; 4) be composed of a majority of DAF civilians; 5) consist of persons of a higher grade than those being considered; 6) use military members in the grade of Colonel or higher, as required; and 7) consider a sufficient number of merit pay candidates to ensure that the merit pay fund is substantial enough to make it possible to reflect differences in performance among the employees competing for a share of the assigned merit pay pool.

The MPRB will meet at the discretion of the convening authority to consider all candidates for merit pay increases. Merit Pay selection folders will be constructed for each recommended employee which will contain the latest performance appraisal workplan and rating form. Selection folders will be sorted by job family as broadly defined by AF Civilian Personnel. Each panel member will review the employee's folder within the limits of specified job families (i.e., all candidates within one job family will be considered by the same panel) using a set of specified criteria. Ratings by each member will be scored, and any deviation of more than five points on total score by any member from any other member will cause the employee's folder to be re-evaluated by the panel members.

Once the board has determined a score for each function, the board

score will be added to the supervisor's rating, and the sum will be multiplied by the percentage weight for that function. Then all the function totals will be summed to determine a merit score. Mathematically, the equation would be:

$$MtSc = \sum_{f=1}^{n} w_f \ (r_f + b_f)$$ where:

a) MtSc is the merit score
b) n is the number of functions
c) w_f is the workplan weight for function f
d) r_f is the supervisor's performance rating on function f
e) b_f is the board score on function f

$$\text{and } b_f = \sum_{c=1}^{k} S_{fc}$$

These scores (r_f, b_f) may each be adjusted by some type of transformation (perhaps specific to a given panel) to control the variance they contribute to the merit score.

The amount of merit pay increase awarded will depend upon the relative place of the candidate's score in the overall ranking of scores. It is anticipated that all GM employees who receive a rating of Fully Successful or higher would receive as a minimum merit pay increase the additional one-half of the comparability pay because their rating indicates that they have performed at or above the expected level. Several methods of determining break points in the overall ranking are contemplated such as, the selection of natural break points in the ranked scores; or the use of a normal distribution percentage as the breakpoint. When all computations have been completed, and merit pay has been assigned to the appropriate individuals, the chairperson of the MPRB will sign off to indicate that the board's decision is the official merit pay determination. After the initial determination, the merit pay list will be distributed as a percentage of the midpoint salary for that grade. The amount which each individual or group of individuals will receive for various positions in the distribution is currently being determined by the Air Force Director of Civilian Personnel.

Conclusions

The conceptual model, as expressed above, is being field tested at several locations within the continental United States. To date, the function approach appears to be working satisfactorily and has been accepted as workable by the employees involved in the field test. Writing standards is causing some difficulty to supervisors and employees, but this was an expected outcome. Supervisory ratings have been collected and are in the process of being scored prior to a mock selection board being established. At least two mock selection boards will be impaneled to run through the selection process, and the results of the boards will be compared to ascertain the reliability of the model. At this point in time, it appears that the appraisal system, as described, can be initiated by the first of October 1980 with the final implementation and the first merit pay awards being made in the first full pay period after the first of October 1981.

INITIAL DEVELOPMENT OF A JOB PERFORMANCE EVALUATION SYSTEM (JPES) FOR AIR FORCE CIVILIAN EMPLOYEES

Thomas W. Watson, John A. Querrieri, and James A. Earles
Air Force Human Resources Laboraotry
Brooks AFB, TX 78235

Abstract

Since 1977 the Air Force Human Resources Laboratory has been involved in the development of a new approach to civilian job performance evaluation. This work was first initiated in response to a 1976 Request for Personnel Research from the Directorate of Civilian Personnel, Hq USAF (RPR 76-40) and has also been guided by the requirements of the Civil Service Reform Act (CSRA) of 1978. The intent is to develop a system which will encourage employee participation in establishing performance standards, effectively differentiate among personnel along a job-performance dimension, and provide performance ratings based on objective criteria. This paper describes initial development of the system through three pre-implementation test phases. Future plans are also discussed.

Since 1977, the Air Force Human Resources Laboratory (AFHRL) has been involved in the development of a new approach to civilian job performance evaluation as one component of a comprehensive civilian appraisal research program. This program, originally initiated in response to a 1976 Request for Personnel Research from the Directorate of Civilian Personnel, HQ USAF (RPR 76-40), has been guided more recently by the requirements of the Civil Service Reform Act (CSRA) of 1978, and a revised RPR which better reflect these requirements.

The need for a new method of job performance evaluation evolved from the inability of current and past methods to differentiate effectively among personnel due to rating inflation. Also, ratings have been criticized by employers, researchers, members of Congress, and the general public for being based primarily on subjective judgments rather than on clearly defined objective performance criteria. For a detailed discourse on the rationale for the job performance research, see Toedt and Ratliff (1978).

Problem

The CSRA of 1978 requires each government agency to develop one or more appraisal systems which will encourage employee participation in establishing performance standards and provide periodic rating of employee performance. The law also requires development of performance standards which permit accurate evaluation on the basis of objective rather than subjective criteria. Performance ratings are to be used as a basis for a variety of personnel actions such as the awarding of quality salary increments or outstanding performance ratings, training, reassignment, promotion, reduction in grade, retraining, and removal.

In addition to fostering subordinate participation and providing objective performance measurement criteria, it is also desirable,

according to the personnel management literature, that performance evaluation systems possess other desirable attributes. They should be relatively immune to rating inflation or system "gaming", provide accurate information concerning work objectives to higher-level management, be acceptable to the work force as well as to management, and be defensible in court. In addition, they should not lessen productivity; but instead, possibly enhance productivity by serving an effective goal-setting and counseling function.

Method

In AFHRL the emphasis has been on taking a research-oriented approach to developing and testing the new job performance evaluation system. Following review of the appraisal literature describing existing approaches to job performance evaluation in government and industry, and consultation with experts in this field, a prototype system was developed considered to be responsive to the requirements of the CSRA of 1978. This prototype system was called Evaluation by Objectives or EBO. More recently, to avoid confusion with the popular Management by Objectives approach (see Odoirne, 1965), the name was changed to Job Performance Evaluation System or JPES. The Air Force Directorate of Civilian Personnel (AFMPK) and the Office of Civilian Personnel Operations (OCPO) collaborated with AFHRL in developing the initial system and in planning the research designed to evaluate the system. As evaluation progressed, various consultants, from both academia and industry, also assisted.

The CSRA of 1978 requires that prototype performance evaluation systems be ready for operational implementation by October 1981. Such time contraints imposed some limitations on research design. For example, rather than implement and evaluate pretests of the system at various locations in a sequential fashion, some overlap between multiple pretests occurred. In addition, the system was continually undergoing refinement (within the limits of CSRA requirements) based on feedback received from the field. This was considered a practical necessity.

Procedure

The system has evolved thus far through a series of three pretests at the following Air Force bases: Andrews, McClellan, and Norton. At each base, participants were given training on how to use the system. This involved training in the development of workplans which reflect the major requirements of each individual's job and specific, objectively defined performance standards. Training also involved informing participants on how performance would be rated based on the objective criteria contained in the workplans.

Training was followed by workplan development in which supervisors and subordinates jointly determined major job requirements and appropriate levels of performance for a variety of job tasks. This was followed by monitoring of job performance and rating of such performance based on the workplans developed. In the Andrews pretest, only one rating period was involved and it was of short duration - two weeks. In the McClellan and Norton pretests, two performance rating periods were involved in order to better determine the effectiveness of the system in terms of both rating inflation and workplan modification across rating periods. Also, in the McClellan and Norton pretests, rating periods were

more realistic than in the Andrews pretest, each spanning approximately one to three months. (Time constraints dictated against even longer rating periods).

In addition to actual field implementation, surveys were developed and administered to participants at each base to determine attitudes toward training and toward the entire system. Face-to-face individual and small-group interviews were also conducted.

The procedures outlined above were implemented at all three pretest sites. However, since the system was constantly evolving, some procedures were unique to specific bases. For example, at Andrews and McClellan AFBs participants were intentionally selected from diverse organizations in order to better determine impact of the system in varying occupational environments. At Norton, participants were selected in larger numbers from two rather different organizations rather than selecting a few subjects from many organizations. This allowed a better assessment of the impact on entire organizations. Also, at Andrews and McClellan AFBs the level of expected performance was anchored or defined at the Marginal level while at Norton, three levels of job performance standards were defined: Marginal, Fully Successful, and Outstanding. At Norton, supervisors and co-workers were requested to rank order workgroup members in terms of their perceived productivity. This aspect was included as a validity check of job performance ratings. Other minor changes also took place as the system evolved at different pretest sites.

An additional pretest of the system is planned at an as yet undetermined site. This pretest will involve testing and evaluation of the promotion potential as well as the job performance evaluation component of the civilian appraisal program.

Pretest Dates

The Andrews pretest was conducted during September and October of 1978; the McClellan pretest, from January through June of 1979; and the Norton pretest, from April through November of 1979.

Subjects

In the Andrews pretest, 66 supervisor/subordinate pairs from diverse organizations participated. In the McClellan pretest, 75 reviewers (second-level supervisors) and 150 worker/supervisor pairs, also from diverse organizations, participated. Sample selection was based on a quintuplet structure whereby each reviewer reviewed the work plan development of two supervisor/subordinate pairs. Persons from the following two organizations participated in the Norton pretest: The 63rd Civil Engineering Squadron, and Detachment 6, 2762 Logistics Squadron. One-hundred seventy-one persons from the former organization participated; 145 from the latter.

In all cases, participants included Air Force civilian employees spanning most grade levels in the GS, WG, WL and WS pay plans, who, for purposes of these tests, served either as job incumbents, supervisors, or reviewers. A job incumbent was defined as someone who had a workplan developed for his or her job. Supervisors were persons providing direct (first-line) supervision, and reviewers were second-line supervisors who reviewed workplans and ratings. Reviewers did not participate at Andrews. Occasionally, military personnel served in a supervisory or

reviewer capacity.

In the Norton pretest, unlike other pretests, workplans were written for the jobs of several enlisted military personnel, as well as for civilians. The intent was to determine the extent to which the new performance evaluation system could be applied to military jobs, even if the new approach was not intended for implementation with this target population.

The target population for which workplans are expected to be developed under the proposed JPES system are persons in the grade of GS 1-13, and all WG, WL, and WS grades. A separate system, the General Manager Appraisal System (or GMAS), is being developed for supervisory and management personnel in the GS 13-15 grades. Another system has been developed for personnel in the Senior Executive Service, and will be referred to as the Senior Executive Appraisal System (SEAS).

Results

Participant reaction and evaluation of the effectiveness of the system, as determined from analysis of workplans, survey data, and through personal interviews, was generally mixed. Some aspects were quite favorably received, while others were frustrating to participants. In some respects, participants were easily able to fulfill the requirements of the system; in others, difficulties were experienced.

Participants experienced some difficulty identifying and defining major job requirements and standards associated with specific job tasks. This was especially true for workers in the jobs which were not product-oriented nor repetitive. It was difficult for some workers and supervisors to define job tasks in terms of standards which were clearly objective and measurable in a practical sense.

It was also discovered that development of effective yet cost-efficient training techniques was a difficult task. Some participants were dissatisfied with the training they had received.

More positively, most participants felt that the concept of the new system was a good one, even if difficulties were experienced putting it into practice. The goals of employee participation and objective performance criteria which would reduce subjective supervisory bias were appreciated.

Rating inflation appeared to be under control. For example, in a range from 500 to 1000, the mean ratings following the first rating phase at Norton were 786.2 for Civil Engineering and 790 for Detachment 6, with standard deviations of 60.1 and 68.5 respectively. Such mean ratings correspond to a fully successful classification which would be expected for most of the work force. Fortunately, when first and second rating phases were compared, mean ratings did not differ significantly.

Participant reaction to the overall system was interesting. As might be expected, some resistance was experienced. However, based on survey data, participant evaluation was more favorable than was indicated during small-group interviews. Perhaps a group dynamic was operating which needs to be considered when drawing conclusions concerning user acceptance. Based on survey response at Andrews, 90% liked the new system, and 94% were satisfied with their workplan. At McClellan and Norton, survey responses were somewhat less favorable. However,

participants at these bases had a more extended and therefore more realistic exposure to the system. Nevertheless, reaction, as measured via survey, was moderately favorable. At McClellan, 68% of the persons for whom workplans were developed considered the new system an improvement over the present system. Likewise, following the first phase of ratings at Norton, 76% of the supervisors felt that the proposed system was better than the present system, or were uncertain, while 24% thought it was worse than the existing system.

Results of the personal interviews were less encouraging. JPES was considered to involve too much paperwork and to be susceptible to "gaming" when operationally implemented. Many participants were concerned about the link between JPES and promotion, and most wanted a normative, standardized approach rather than an idiographic, ipsative approach. Many of the potential problems which surfaced during these discussions appear not to be solely directed toward JPES, but reflect federal-service-wide concerns which have been addressed in recent issues of the Federal Times.

Conclusions and Recommendations

Although JPES has undergone continual modification in what appears to be a positive direction, it still appears to be in need of further refinement and evaluation. Attitudes and opinions, as elicited through surveys and interviews have been a fruitful source for system refinement.

As with any new system, some resistance to change has been experienced, and the system has been criticized by participants who experienced difficulties in writing workplans and who were frustrated by the paperwork involved. However, most persons have been receptive to the goals of providing objective criteria for performance evaluation, facilitating communication between supervisors and subordinates, and participation by those impacted by the system.

Due to the mixed participant reaction, certain aspects will need to be modified to increase user acceptance while continuing to meet CSRA requirements. Also, the relationship of the JPES component to the promotion potential component will need to be clarified, and the system will need to be marketed more effectively

Training is a particularly critical area in which improvement appears to be needed. By providing more intensive and effective training, many of the problems experienced during the pretests could be resolved.

Additional analysis of existing pretest data is in order. Also, another more rigorously controlled pretest which includes system modifications based on input received to date, and which includes other appraisal system components, is planned. This should provide additional insight into the optimum configuration of a performance evaluation system.

References

Odiorne, G. S., Management by Objectives. Belmont, California; Pitman, 1965.

Toedt, D. C., & Ratliff, F. R. System psychology applied to performance assessment of USAF civilian employees. Paper presented at the Sixth Annual Psychology in the Department of Defense Symposium, USAF Academy, Department of Behavioral Sciences and Leadership, Colorado Springs, Colorado, 20-22 April 1978.

RATER ACCURACY IN PREDICTIONS OF ABILITY

Capt William H. Cummings
Air Force Human Resources Laboratory
Brooks Air Force Base, Texas

Abstract

An extensive test of the civilian Promotion Potential Appraisal System provided an excellent opportunity to examine a number of questions concerning rater accuracy. In this study, 265 civilian employees at McClellan Air Force Base took a battery of aptitude tests and received ratings from three individuals (supervisor, co-worker, and supervisor's associate). Rater accuracy, as determined by rated ability versus actual test scores, generalized significantly ($p < .01$) across the different test dimensions. This effect replicated previous findings and extended them to the operational setting. Furthermore, there was significant generalization even in the absence of any correlation between test dimensions. Raters accurately ($p < .01$) predicted the employees' test scores; and supervisors, co-workers, and associates were about equally accurate as raters.

Most people, it seems, feel that they can accurately evaluate other people that they know well. This is especially true when the evaluation is an important one, such as a job performance rating, which may have significant monetary implications. An important question, though, is whether some people are more accurate than others in making these ratings. This is a question with important research and operational implications. If there are indeed accurate-rater "types," and if we can identify them and their characteristics, then we might be able to improve the precision of the rating instruments we use. We might even be able to select supervisors and reporting officials partially on the basis of their demonstrated accuracy as raters.

Two studies from the Human Resources Laboratory have already provided some evidence for rater accuracy as a generalized ability. Mullins and Force (1962) found that raters who were the most accurate at estimating vocabulary test scores were also the most accurate at making ratings of carefulness. Mullins, Seidling, Wilbourn, and Earles (1979) later replicated this finding and extended it to different traits.

These findings lead to a number of relevant questions. The primary question at this point is whether these findings are indeed due to a trait-like property of rater accuracy or, instead, to the underlying correlation between the dimensions being evaluated. For example, if the dimensions being evaluated are positively correlated, the so-called accurate rater may not be making an accurate, independent assessment of the different traits. He may simply know the subject's ability on one dimension and infer the subject's ability on the other dimensions. In this case, he would get credit for being accurate on both dimensions. Such a phenomenon would be consistent with recent findings indicating that most raters are able to provide only a very limited amount of

useful information on the people they are rating.

Another question is whether the previous findings can be extended to the operational setting. Closely related is the question of whether people in different rater roles (e.g., supervisor, co-worker) can make more accurate ratings. With numerous studies touting the virtues of peer ratings (see, e.g., Dunnette & Borman, 1979), the question arises whether individuals in different rater roles differ in their accuracy.

A recent test of the proposed civilian Promotion Potential Appraisal System (PPAS) provided an excellent opportunity to examine these questions. Civilian employees at McClellan Air Force Base took a series of tests. Later, each worker was rated on the tested characteristics by three individuals: his supervisor, a co-worker, and an associate of his supervisor. By comparing the ratings with the test scores, we were able to compute a series of rater accuracy scores on each rater and further investigate rater accuracy as a trait.

Method

Subjects were 265 nonsupervisory civilian employees at McClellan AFB. A supervisor and two workers from the same shop were selected so that the two workers could serve as co-workers for rating each other. A co-worker of each supervisor (supervisor's associate) provided additional ratings. This setup provided for three raters for each worker: the supervisor, the supervisor's associate, and the matching co-worker.

Each worker first completed a Civilian Personnel Examination (CPE), a Score-Checking Test (SCT), and a demographic questionnaire. The CPE is a modified version of the Airman Qualifying Examination-Form J (Vitola, Massey, & Wilbourn, 1971). The CPE is designed to assess the worker's abilities in 10 areas that emphasize blue-collar and clerical skills. The SCT is a carefulness test. It requires the examinee to check two sets of supposedly identical numbers for transcription errors.

Approximately 1 month after testing, each worker received ratings from the three different raters. These ratings provided 101 different measures from each rater. Certain portions of the rating booklet were specifically designed to tap rater accuracy. One section required the rater to evaluate the worker on seven general abilities (arithmetic computation, word knowledge, data interpretation, electrical information, shop procedures, mechanical skill, and carefulness). These abilities match seven of the CPE subtests and the SCT. Another section of the booklet required the rater to rank-order the same seven general abilities in order of the worker's best qualities. Yet another section provided an estimate of the worker's exact score on the word knowledge (vocabulary test) portion of the CPE.

Rater accuracy scores were computed as follows:

$$\text{Rater Accuracy} = \text{absolute} \left| z_{(\text{normative rating})} - z_{(\text{test score})} \right|$$

The conversion to z-scores eliminates problems due to scaling differences between the normative measure and the corresponding test score. Use of the absolute value of the difference between these

Table 1
Intercorrelations of Rater Accuracy Measures
(n = 660[a])

	2	3	4	5	6	7	8	9
1. Arith Comp Acc	.21	.10	.11	.14	.20	.21	.04	-.11
2. Data Interp Acc		.19	.21	.19	.28	.23	.18	-.07
3. Elec Info Acc			.39	.36	.21	.03	.14	-.02
4. Gen Mech Acc				.51	.16	.20	.10	-.10
5. Shop Prac Acc					.13	.18	.08	-.09
6. Word Know Acc						.21	.55	-.04
7. Carefulness Acc							.14	-.11
8. Voc Test Acc								-.09
9. Rank Ability Acc								-

Note. .01 Significance Level = .094. Negative correlations between variable 9 and other variables reflect a positive relationship between the different accuracy dimensions, since the scales are reversed.

[a] n = 595 for Rank Ability Accuracy, due to data lost in the rank-ordering procedure. The .01 Significance Level = .098.

scores results in a value that decreases as rater accuracy increases, regardless of whether the rater overestimates or underestimates the worker's ability. Thus, smaller values on these rater accuracy indices reflect increased rater ability.

A final accuracy measure was based on the correlation between the rater's rank-ordering of the worker's abilities and the worker's actual test scores. This measure, Rank Ability Accuracy, shows how well the rater knows what the worker's best--and worst--traits are.

There are, then, nine rater accuracy measures: the seven normative-based accuracy measures (Arithmetic Computation Accuracy, Data Interpretation Accuracy, Electronic Information Accuracy, General Mechanical Accuracy, Shop Practices Accuracy, Word Knowledge Accuracy, and Carefulness Accuracy), the exact-estimate-based accuracy measure (Vocabulary Test Accuracy), and the rank-order accuracy measure (Rank Ability Accuracy).

Results and Discussion

Complete rating sets were received from 230 supervisors, 198 supervisor's associates, and 235 co-workers. The data were collapsed across the three rater roles so that only one data set (n = 663) is presented here, instead of three. However, the data were also analyzed separately for each rater role so that deviations from the overall pattern could be noted.

Table 1 presents the intercorrelations of the nine different rater accuracy measures. (For ease of interpretation, due to the large number of coefficients involved, the .01 significance level was used, rather than the more traditional .05 level.) It is obvious that the rater accuracy measures are significantly intercorrelated, though the relationships are hardly overwhelming. Twenty-eight of the 36 possible

Table 2
Intercorrelations of CPE and SCT Measures
(n = 265)

	2	3	4	5	6	7
1. Arith Comp	.41	.19	.16	.14	.42	.53
2. Data Interp		.40	.36	.35	.52	.50
3 Elec Info			.83	.76	.36	.11
4. Gen Mech				.83	.27	.07
5. Shop Prac					.23	.05
6. Word Know						.46
7. SCT						-

Note. .01 Significance Level = .144.

relationships are significant, and all are in the predicted direction. This finding replicates the previous rater accuracy results of a consistent but relatively weak relation between different dimensions of rater accuracy.

The crucial question, however, is whether or not this trend can be attributed to the intercorrelations between the aptitude tests. Table 2 shows clearly that the tests are, in fact, moderately to powerfully intercorrelated. Moreover, the correlation between the coefficient for a given pair of rater accuracy measures and the coefficient for the corresponding pair of test measures is quite high: $r(19) = .91$, $p < .005$. Thus, a good deal of the "rater accuracy effect" could certainly be attributed to the correlation between tests. Knowing a person's ability on one dimension certainly helps in predicting his ability on another closely related dimension.

However, a more analytical version of this question is whether there is a rater accuracy effect even after the correlation between tests is removed. Table 2 shows that four pairs of tests are not significantly correlated (tests 1 and 5, 3 and 7, 4 and 7, and 5 and 7). Table 1 indicates that the rater accuracy measures are significantly intercorrelated for three of these four test pairs. Thus, even in the absence of any test intercorrelation, there is still some rater accuracy effect.

The Rank Ability Accuracy measure (variable 9) provides further evidence on this issue. Three of the eight relationships with the other rater accuracy measures are significant at the .01 level, and all of the correlations are in the predicted direction. This finding is particularly interesting because the correlation between tests cannot account for the correlations with this particular measure. Moreover, these results extend the generality of the rater accuracy effect. Not only can the accurate rater evaluate the strength of his worker's abilities in a more-or-less absolute sense; he can also evaluate the relative strength of these abilities.

The second major question addressed here is whether raters in different roles (supervisor, supervisor's associate, co-worker) provide more accurate ratings. The answer, at least in terms of the present data, is simple: no. There were no significant ($p < .05$ by one-way

analysis of variance) differences in accuracy scores between supervisors, supervisor's associates, and co-workers. Supervisors had the most accurate scores on four measures, supervisor's associates on four measures, and co-workers on one measure.

The lack of any detectable differences in rater accuracy between rater roles is somewhat surprising. It seemed likely that the supervisor would be more accurate because of his history of observing and rating the worker on job-relevant dimensions. Alternatively, a case could be made for the co-worker because of numerous opportunities to observe the worker. However, the associate was (for all practical purposes) as accurate as the supervisor and the co-worker. This phenomenon deserves further study, especially in the context of previous findings demonstrating the superiority of peer ratings. If our results can be replicated, they may have important implications for the use of different rater roles in job performance evaluations.

It should be noted that in terms of absolute rater accuracy, all three types of raters were at least reasonably accurate on the measures used here. For supervisors, the median correlation between the raw ability rating and the worker's corresponding test score was .38 (range: .09 to .57, with 8 of 9 $\underline{r}$'s significant at .01). For supervisor's associates, the median correlation was also .38 (range: .14 to .61, with 8 of 9 $\underline{r}$'s significant). For co-workers, the median correlation was .30 (range: .06 to .59, with 8 of 9 $\underline{r}$'s significant). Accuracy of carefulness ratings was uniformly low ($\bar{r}$ = .09, .14, and .06, respectively, for the three roles). On the other hand, accuracy in predicting the shop-related test scores (Elec Info, Gen Mech, and Shop Prac) was generally high (range: .30 to .61). These accuracy scores indicate that the failure to detect any significant differences between rater roles is not due to overall rater inaccuracy or imprecision of the measures. Our raters were making good, valid ratings. It appears that, at least in this study, there were simply no reliable differences between different rater types on some fairly concrete dimensions.

References

Dunnette, M.D., & Borman, W.C. Personnel selection and classification systems. In M.R. Rosenzweig & L.W. Porter (Eds.), Annual review of psychology (Vol 30). Palo Alto CA: Annual Reviews, Inc., 1979.

Mullins, C.J., & Force, R.C. Rater accuracy as a generalized ability. Journal of Applied Psychology, 1962, 46, 191-193.

Mullins, C.J., Seidling, K., Wilbourn, J., & Earles, J.A. Rater accuracy study (AFHRL-TR-78-89). Brooks AFB TX: Air Force Human Resources Laboratory, February 1979.

Vitola, B.M., Massey, I.H., & Wilbourn, J.M. Development and standardization of the airman qualifying examination - Form J (AFHRL-TR-71-28). Brooks AFB TX: Air Force Human Resources Laboratory, May 1971.

The Effect on Leniency of Justifying Performance Ratings to a Supervior

Phillip J. Decker
University of Missouri-St. Louis

and

Edwin T. Cornelius, III
The Ohio State University

Abstract

This study was undertaken to examine the effect of a perceived requirement to justify ratings to a supervisor on the amount of leniency in the resultant ratings in a rating procedure where all ratings were justified to the ratee. Eighteen NROTC cadet officers rated 73 midshipmen in the normal quarterly evaluation; one half perceiving the justification to superior officer requirement, the other half not. The experimental condition ratings were more lenient than the control ratings, contrary to expectation. Questionnaire results indicated that rater self-esteem was an important factor - raters who were to justify their ratings indicated that low ratings would result in lowered self-esteem. It was suggested that the military setting was critical to the study results.

Few areas of industrial psychology have recieved more attention than the use of supervisory ratings of job performance as an element of criteria data. Yet, despite the attention, this means of performance appraisal has been plagued by problems, rater errors being the most commonly cited obstacle to accurate ratings. The use of supervisory ratings rests on the assumption that the supervisor is a good instrument of quantitative observation and that (s)he is capable of objectively scoring the observations (Guilford, 1954). One of the major obstacles to accurate ratings has been leniency (Guion, 1965) and especially so in military settings (Cornelius and Hakel, Note 1).

Cascio (1978) has suggested that leniency can be controlled in two ways: 1) by allocating ratings into a forced normal distrubiton, or 2) by reducing the ambiguity in the rating scales themselves. However, raters tend to resist forced distributions (Sharon and Bartlett, 1969); a lenient distribution may, in fact, be correct (Cascio and Valenzi, 1977), and past efforts at reducing leniency through changing scale format have been far from successful (Landy and Farr, 1979). Because of these results, recent studies have attempted to look at other variables affecting leniency, such as organizational roles (Borman, 1974), participation of raters in scale construction (Freidman and Cornelius, 1976), and training of raters (Warmke and Billings, 1979).

Much of the reasoning for the shift of focus shown in the recent research on leniency may be explained by viewing leniency primarily as a function of rater motivation (Decker, Note 2). Both Glickman (1955) and Bass (1956) have suggested that leniency can be caused by the rater's appraisal of the consequences to himself and the ratee of producing ratings at a given level. A review of the merit rating literature reveals several studies in which instructions were given to raters in such a way that they perceived different consequences resulting from rating level. Stockford and Bissel (1949) found that supervisors who perceived that they were identified to the ratees and perceived a requirement to "explain" their ratings to their subordinates rated them more leniently than when they could withhold the ratings from subor-

dinates. Taylor and Wherry (1951) found that ratings made for "experimental" reasons had a lower mean level than did ratings made for "administrative" reasons. Finally, Sharon and Bartlett (1969) examined the effects of four experimental conditions on leniency: 1) evaluation - in which the the rater was identified but the ratings were used by another person to evaluate the ratee, 2) identification -in which the rater put his name on the rating form but the ratings were used for research purposes only, 3) justification - in which the rater perceived identification and that (s)he would have to justify the ratings to the ratee in a face-to-face meeting, and 4) a control condition. These researchers found that the evaluation and justification condition raters were more lenient than raters in the other two conditions.

Wherry (Note 3) has suggested the requirement that ratings be seen by both the ratee and rater's superivsor will tend to cancel out the effects of showing the ratings to only one or the other. Landy and Farr (1979) indicate that this procdure is commonly used in industry, but has not been empirically tested. The purpose of this study was to examine this procedure in a military setting where leniency has been a reoccurring problem. If the perceived requirement to justify ratings to ratees affects motivation and hence behavior, it is logical to expect a perceived requirement to justify ratings to the rater's supervisor to do the same. If a superivor expects non-lenient ratings in his/her department, defines leniency, communicates this to his/her subordinates who are rating, and requires raters to justify ratings to him/her, the result should be less lenient ratings. If supervisor expectations are important to the raters, it seems reasonable for the raters in this situation to avoid lenient ratings in order to please their supervisor and minimize undesirable consequences from the supervisor. In indirect support, Sussman (1974) has shown that in upward communication, message senders in organizational hierarchies structure messages to strengthen their own position.

Method

Subjects and setting. The raters were 18 undergraduate students participating in NROTC at Ohio State University. The raters were members of two companies and companies were randomly assigned to the experimental and control conditions. Companies were matched on race, sex, educational level, and service orientation (Marine vs. Navy). The 18 midshipmen officers rated 73 cadets for the regularly scheduled quarterly evaluation with the existing rating form used in NROTC (NAVEDTRA 16 10/1, 6-74). This form utilized 10 point scales on five dimensions: professional performance, military behavior, leadership, supervisory ability, military appearance, and overall. The leadership dimension scores were not analyzed in this study because not all cadets were evaluated on it.

When the ratings were completed by all the raters, a post-experimental questionnaire was distributed which was designed to elicit the motivational process leading to the level of ratings and manipulation induction checks. The motivation process section was based on a motivation questionnaire used by Lawler and Suttle (1973), and incorporated several modifications suggested by Mitchell (1974). Confidentiality of responses was insured.

Fifteen valences/expectancies were included: ratee arguing, increases ratee performance, ratee likes me, decreases ratee performance, poor data to NROTC, ratee gets a "big head," poor data to ratee, good data to NROTC, increases my self-esteem, good data to ratee, descreases my self-esteem, ratee believes promotion truly earned, keep ratee from promotion, make supervisor like me, and make supervisor dislike me.

Procedure. The procedure followed was the standard practice followed in past evaluation in the NROTC unit, except for the instructions necessary for the experi-

mental manipulation and the completion of a post-experimental questionnaire. This procdure included justifying ratings to the ratee. The experimental condition was created by verbal instructions from the NROTC Executive Office (EO) that indicated 1) raters would be required to justify their ratings to him and he was greatly concerned about the problem of leniency. Leniency was defined as predominately high ratings within rater and the term "rating inflation" was used in the instructions. The raters in the experimental condition (Company "A") were also told by the EO that not all raters would be required to justify their ratings to him that particular quarter because he did not have the time to meet with each rater. The raters in the control condition (Company "B") were told that some raters would be meeting with the EO but that they would not be. This statement was used because it was thought that rater interaction may cause diffusion of treatment (Cook and Campbell, 1979).

Leniency was assesed by comparing mean ratings across rating condition. Because of the complexity of the military command in the NROTC unit, each rater rated a different number of midshipmen and in two cases (the midshipmen company commanders) rated the same cadets as other raters. Consequently, for the ratings of each rater, means were computed by rating scale dimension. The motivational questionnaire was filled out by the raters on their own time after all rating was completed.

Results

The post-experimental induction checks revealed that all raters perceived a requirement to justify ratings to the ratees and that experimental subjects perceived a requirement to justify ratings to the EO where control condition raters did not. The induction check item responses also revealed that all raters perceived the EO was mildly concerned about leniency in ratings; there were no significant difference across conditions.

Table 1 shows the means and standard deviations of the rater's scores on all the midshipmen (s)he rated. In order to test the effects of justifying ratings to a superivor on leniency, a multivariate analysis of variance was performed using the experimental conditions as the independent variables. This analysis reached a level of significance (p .05) using the Wilks' Lamda criterion. Overall rating means and standardized discriminant function coefficients are reported in Table 2. As Table 2 shows, the overall rating score means for the experimental group were, for the most part, higher than those of the control group. Perceptions of a requirement to justify ratings to a supervisor caused leniency as defined herein.

The post-experimental questionnaire was used to calculate effort-to-rate-low and effort-to-rate-high scores. These data are presented in Table 3. Within company comparisons parralleled the actual rating scores by condition. Further analysis revealed that the variance of these scores was primarily explained by rater self-esteem; the experimental condition raters felt that low ratings would lead to decreased self-esteem (and that high ratings would lead to increased self-esteem). The opposite views were held by control group raters. There were some slight valence differences across conditions, but the items where differences were found did not add to the variance of the effort-to-rate scores. Diverse items focusing on ratee performance, ratee and supervisor's feeling toward the rater, organizational effectiveness, and feedback to the ratee did not contribute to the effort-to-rate scores.

Discussion

The raters in the experimental condition rated leniently because of the perceived consequences of their ratings. This was not related to interpersonal relations with the ratee (because this was a design constant) nor directly to interpersonal relations with the supervisor reviewing the ratings (questionnaire results). Bass (1956) suggested,

among other things, that raters rate leniently because they think "whoever associates with me is meritorious, therefore I am also (p. 359.)" This feeling seems to be enhanced in a military setting where the raters perceive a requirement to justify their ratings to a superior officer. Consequently, in military settings, at least, the effects of justifying ratings to ratees and supervisors do no cancel out as Wherry (Note 3) suggested; but increase the liklihood of leniency.

Many ratings in industry are justified either directly or indirectly to a supervisor; therefore, the results of this study, though admittedly influenced by the ongoing dynamics of the military setting, are an indication that the effects of justifying ratings upward may be a critical factor in rater motivation and the resulting psychometric properties of the ratings. It is fairly clear that consequences of a particular rating condition or procedure must be thoroughly understood. This endeavor is clearly in its infancy.

Reference Notes

1. Cornelius, E.T., III and Hakel, M.D. A study to develop an improved enlisted enlisted performance enlisted system for the U.S. Cost Guard. Final Report, Grant # DOT - CG - 62625 - A, The Ohio State University, 1977.
2. Decker, P.J. The effect on leniency of justifying performance ratings to a supervisor. An unpublished Master's thesis, The Ohio State University, 1977.
3. Wherry, R.J. The control of bias in rating: A theory of rating. PRB Report No. 922, Contract # DA - 49 - 083 OSA 69, Department of the Army, 1952.

References

Bass, B.M. Reducing leniency in ratings. Personel Psychology, 1956, 2, 359-369.

Borman, W.C. The rating of individuals in organizations: An alternative approach. Organizational Behavior and Human Performance, 1974, 12, 105-124.

Casico, W.F. Applied Psychology in Personnel Management. Reston: Reston, Virginia, 1978.

Cascio, W.F. and Valenzi, E.R. Behaviorally anchored rating scales: Effects of eduction and job experience of raters and ratees. Journal of Applied Psychology, 1977, 62, 278-282.

Cook, T.D. and Campbell, D.T. Quasi-Experimentation: Design and Analysis Issues for Field Settings. Rand McNally: Chicago, 1979.

Friedman, B.A. and Cornelius, E.T., III. Effect of rater participation in scale construction on the psychometric characteristics of 2 rating scale formats. Journal of Applied Psychology, 1976, 61, 210-216.

Glickman, A.S. Effects of negatively shewed ratings on motivations of the rated. Personnel Psychology, 1955, 8, 39-47.

Guilford, J.P. Rating methods, Psychometric Methods. New York, McGraw-Hill, 1954.

Guion, R.M. Personnel Testing. McGraw-Hill: New York, 1965.

Landy, F.J. and Farr, J.L. Performance rating. In press, Psychological Bulletin, 1979.

Lawler, E.E. and Suttle, J.L. Expectancy theory and job behavior. Organizational Behavior and Human Performance, 1973, 9, 482-503.

Mitchell, T.R. Expectancy models of job satisfaction, occupational preference and effort. Psychological Bulletin, 1974, 81, 1053-107.

Sharon, A.T. and Bartlett, C.J. Effect of instructional conditions in producing leniency on two types of rating scales. Personnel Psychology, 1969, 22, 251-263.

Stockford, L. and Bissel, H.W. Factors involved in establishing a merit rating scale. Personnel, 1949, 26, 94-118.

Sussman, L. Perceived message distortion or you can fool some of the supervisors some of the time... Personnel Journal, 1974, , 679-682.

Taylor, E.K. and Wherry, R.J. A study of leniency in two rating systems. Personnel Psychology, 1951, 4, 39-47.

Warmke, D.L. and Billings, R.S. Comparison of training methods for improving the psychometric quality of experimental and administrative performance ratings. Journal of Applied Psychology, 1979, 64, 124-131.

Table 1

Means and Standard Deviations of Ratings by Dimension

Experimental Condition Raters (Ratee N = 38)

	1		2		3		4		5		6		7		8		9	
Dimension	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Professional Performance	6.15	1.21	7.33	1.04	6.72	1.12	6.75	1.50	6.80	1.64	7.50	1.00	6.00	0.70	6.00	0.10	7.50	1.29
Military Behavior	6.75	1.14	7.33	1.11	6.89	0.67	6.89	1.26	7.25	1.09	7.75	0.96	6.00	0.71	6.40	0.55	8.75	0.96
Military Appearance	7.20	1.67	7.47	1.55	7.06	1.16	7.00	1.82	7.60	1.14	7.50	1.73	6.40	0.55	6.00	8.50	0.00	1.29
Overall	6.60	1.29	7.40	1.29	6.83	0.79	6.75	1.50	7.00	1.41	7.25	1.26	6.20	0.45	6.20	0.45	8.50	1.29

Control Condition Raters (Ratee N = 45)

	1		2		3		4		5		6		7		8		9	
Dimension	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Professional Performance	6.97	1.36	7.20	1.61	7.20	1.57	5.20	0.45	7.75	0.96	6.33	2.08	7.33	0.58	7.50	0.58	5.50	1.29
Military Behavior	7.02	1.54	7.20	1.66	7.20	1.08	5.40	0.55	7.50	0.58	6.67	1.53	8.33	0.58	7.75	0.96	5.50	1.92
Military Appearance	6.71	1.50	7.13	1.89	6.71	1.58	5.40	1.14	7.75	1.50	5.33	2.52	6.33	0.58	7.75	2.06	6.00	1.56
Overall	6.71	1.56	6.93	1.71	6.80	1.57	5.60	0.89	8.00	0.82	5.67	2.08	6.33	0.58	8.00	1.15	6.25	0.96

Table 2

Overall Means, Standard Deviations, and Discriminant Function Coefficients For Testing Rating Dimension Score Differences Between Conditions

Dimension	Overall Mean	SD	Overall Mean	SD	Discriminant Function Coefficients
Professional Performance	6.75	.61	6.78	.90	1.56
Military Behavior	7.15	.80	6.95	.97	-0.77
Military Appearance	7.19	.72	6.57	.89	-3.44
Overall	6.97	.87	6.69	.87	2.44

Note: Although none of the univariate tests were statistically significant, the linear combination of four variables defined by the discriminant weights seperated the two groups.

Table 3

Means, Standard Deviations, and Ranges of Company Effort-to-Rate Scores

	Effort-to-Rate Low			Effort-to-Rate High		
Company	Mean	SD	Range	Mean	SD	Range
A (experimental)	96.56	117.63	352.5	113.67	117.05	310.0
B (control)	142.78	107.92	307.5	126.44	95.06	272.5

Type I and Type II Errors in Performance Evaluation

Frederick H. Steinheiser
Army kesearch Institute
Alexandria, VA 22333

Abstract

In a classification problem, there will invariably be some false positives and false negatives. For example, in a criterion-referenced test following a training program, some of the students who fail the test may nonetheless be 'true masters,' and some who pass may be 'true nonmasters.' 237 Military Police students took 3 replications (240 shots) of a handgun marksmanship test. Type I and II error rates were each found to be about 12%. Instructors viewed a false positive to be about 6 times as serious as a false negative. ANOVA was conducted to show how different sampling assumptions might mislead the decision maker who does not critically interpret the summary statistics.

Criterion-referenced measurement provides test data which are interpreted in terms of an examinee's ability to achieve an objective or to perform a task. Decisions are based upon the performance scores. Often the decision-making process is a simple dichotomy: students either pass or fail, they are either 'masters' or 'nonmasters,' they are either promoted to the next unit of instruction or are recycled for remedial instruction.

Unfortunately, even very good criterion-referenced tests are not error-free. Items may not adequately reflect the objectives or tasks for which the test was designed, leading to problems of test validity. Human performance is variable, due to fatigue, even for highly skilled tasks. Lapses of attention can degrade skilled performance. Lucky guesses can erroneously inflate a test score. The problem is to assess the error rate in the testing (or measurement) process itself.

The purpose of this experiment was to conduct several replications of a 'hands-on' criterion-referenced test in order to assess Type I and Type II misclassifications: 'masters' incorrectly classified as 'nonmasters' and vice versa. The relative subjective cost of each error type was also estimated. The second aim of the experiment was to determine the effect of different sampling assumptions on summary statistics as computed by analysis of variance (ANOVA). It was reasoned that different, but equally legitimate assumptions could lead to different conclusions about the same set of data.

Method

Subjects. 237 students from the U.S. Army Military Police School participated. Each student had previously passed a handgun marksmanship training course by getting at least 35 hits out of 50 shots at stationary silhouette targets. All students were male, with an age range of 19 to 22 years.

Criterion-referenced test. The Military Police Firearms Qualification Course (MPFQC) is used to certify students in .45 caliber handgun marksmanship. The test must be passed in order to graduate from the M.P.

School, and is given immediately after training. The test normally consists of 50 rounds fired from 8 different conditions, called 'tables' by the School. The tables differ in distance to the target (35 m. to 7 m.) and position (standing, kneeling, crouching) of the shooter. For purposes of this experiment, each student shot 10 rounds for each table, and did this on 3 separate occasions.

Procedure. The total of 240 shots (or 'test items' so to speak) were fired in 3 replications of 80 shots each. The first was fired in the morning, the second in the afternoon, and the third in the morning of the following day. The 10 shots for each table were subdivided into 2 groups of 5 shots each. After taking the 5 shots, the student reloaded his weapon, scores were recorded, and the holes in the target were covered with black tape to prevent feedback. Groups of 20 to 25 students took the test each week; the entire data collection effort required 10 weeks.

Results

Misclassification analysis. The criterion score for passing was set at 70%, in accordance with M.P. School policy. A total of 63 students (27%) failed to get at least 56 hits for a passing score out of the first 80 shots. A total of 58 students (25%) failed to get at least 168 hits for a passing score on the entire 3 replication 240 item test.

In order to estimate misclassifications, the results from the first replication were compared, student by student, to the complete 240 item test. That is, some students who got at least 56 out of the first 80 shots as hits may have gotten fewer than 168 hits out of the entire 240 shot test; and vice versa.

Table 1 shows that 11% of the students who passed the first 80 item test failed to pass the entire 240 item thrice replicated test. This is the false positive rate. Table 1 also shows that 14% who would have been classified as nonmasters on the basis of the 80 item test were incorrectly classified, since they managed to get at least 168 hits on the entire 240 item test. Thus, there is an approximate 25% misclassification rate, if only scores from the first test replication of 80 items were used as the data for classifying the mastery status of a student.

TABLE 1

False Positive and False Negative Misclassifications

		Decision Based Upon 240 Item Test: Pass	Fail
Decision Based Upon 80 Item Test:	Pass	89%	11%
	Fail	14%	76%

Subjective significance of misclassifications. Six M.P. instructors were individually interviewed, in order to elicit expert judgment in evaluating the relative seriousness of a false positive vs a false negative. (They were not told what the empirical rates of Table 1 were.) Each instructor was asked: "Given that "x" number of unqualified students are going to be incorrectly passed (as false positives), then how bad a mistake do you think that is when compared to failing some number ("y") of qualified students?" "x" took on values from 1 to 4, and the instructor merely had to supply his "y" value, so that "x" false positives would be

subjectively comparable in seriousness to the "y" false negatives.

The average perceived seriousness of a false positive was found to be subjectively equivalent to 6 false negatives. That is, the problems associated with failing 6 truly qualified students (in terms of decreased morale, depletion of the manpower pool, additional time and expense to train and test 6 new students) were perceived as comparable to the different set of problem associated with erroneously passing an unqualified student (e.g., the likelihood that this student could not use his weapon appropriately when necessary and thereby inflict unwanted damage/injury).

Analysis of variance. The four factors in this completely crossed experimental design are: 237 subjects x 3 replications of the 30 shot test x 2 'tables' x 2 shot groups of 5 shots each. The following ANOVA testable hypotheses were proposed: (1) most of the variance should be accounted for by the 'subjects' factor if the test is separating masters from nonmasters; (2) large variance due to 'replications' would indicate either a learning or fatigue effect; (3) large variance due to 'tables' would indicate non-homogeneous tables or test items; (4) large variance due to shot groups would indicate a lack of stability in the scores.

Three ANOVAs were conducted using different sampling assumptions: (1) 'subjects' was the only randomly sampled factor; (2) all factors except for 'tables' were randomly sampled from larger populations; (3) all 4 factors were randomly sampled. Highlights of these analyses are presented in Table 2; complete details are in a report by Steinheiser and Epstein (1973), which also has derivations for d.f. and quasi F-ratios.

TABLE 2

Changes in F-Ratio as a Function of Sampling Assumptions

Source	F(1)	F(2)	F(3)
A (Subjects)	---	3.93 $p < .001$	2.54 $p < .01$
B (Replications)	14.18 $p < .001$	12.55 $p < .001$	4.71 $p < .025$
C (Tables)	365.84 $p < .001$	79111 $p < .001$	79.11 $p < .001$
D (Shot Group)	7.33 $p < .001$	5.96 $p < .025$	2.26 n.s.

Is there a statistically significant effect due to 'replications' and to 'shot groups?' As fixed factors (no desire to generalize to additional lanes or more replications), then the answer is "yes." But if they were presumably sampled from larger populations, then the answer is "no," or at least not as statistically significant.

Magnitude of effect summary statistic. The 'magnitude of effect' (sometimes called the 'proportion of variance accounted for') is a summary statistic which measures the degree of association between an independent and dependent variable. This index is less sensitive to sample size than is the F-ratio, does not imply probabilistic hypothesis testing as does F, and should be considered as a source of valuable supplementary information following ANOVA and F ratio computations. Rules for derivation may be found in papers by Dodd and Schultz (1973), and Steinheiser and Epstein (1973). Results for the main effects in this experiment are presented in Table 3.

The largest effect, other than random error, was due to the 'tables' factor, which captured from 10% to 20% of the variability in scores,

TABLE 3

Magnitudes of Main Effects Under Different Sampling Assumptions

Source	A Random, B,C,D Fixed	A,B,D Random, C Fixed	A,B,C,D Random
A (Subjects)	.0852	.1027	.1030
B (Replications)	.0027	.0041	.0042
C (Tables)	.1643	.2454	.2631
D (Shot Groups)	.0004	.0006	.0005

depending upon the sampling assumptions. The effect due to 'subjects,' reflecting individual differences, attained a value of about 10%. The effect due to 'replications' was about .4%, whereas this factor had attained statistical significance according to the F-ratio in Table 2. A similar relationship holds for the 'shot group' factor.

The contrast between F-ratio and magnitude of effect statistics emerges most clearly in some of the interaction terms. Values for some are shown in Table 4, with the columns corresponding to the sampling assumptions of Table 3.

TABLE 4

Statistical Significance and Magnitude of Effect for Interaction Terms

Source	p	M.E.	p	M.E.	p	M.E.
AC	.001	.101	.001	.054	.001	.057
AB	untestable	.049	.001	.044	.001	.037
BC	.001	.002	.01	.003	.01	.004
ABC	untestable	.182	.001	.077	.001	.081

There is contrasting information conveyed by the summary statistics in Table 4. Based upon the F-ratio, these interactions have all reached rather high levels of statistical significance in rejecting the null hypothesis of 'no effect' due to that interaction. But the magnitude of effect index suggests that each interaction accounts for relatively little of the overall variability in the data.

Discussion

Theoretical implications. Criterion-referenced measurement attempts to develop a 'quality control' model for personnel performance evaluation. Those examinees who pass a test are assumed to be masters. Yet realistically, error-free classification is not possible. Several types of mathematical models have been developed in order to reduce classification error. Space limitation precludes explication of the models here, but Epstein (1979) presents a cogent summary, and also shows how they were applied to the current MPFQC data base. Epstein found that there were few differences between the models in terms of the amount of decision-making error, the predictability of the magnitude or direction of the decision error, or in the accuracy of true score estimates based upon the observed test scores.

Relative subjective costs of Type I vs Type II errors were readily obtained from a group of content experts. The cost units are not necessarily monetary. Since false positives were perceived as being much more costly than false negatives, the implication is that the

criterion for passing should be made stricter. As a result, there would be fewer false positives but more false negatives.

The results from the analysis of variance show that, for a powerful test with a large sample size, the selection of the model based upon sampling assumptions can have an impact on the summary statistics. For the MPFQC, it would seem logical to consider the 8 'tables' as a fixed effect, since this experiment was really a random sample in time, of the infinite number of times that students could actually be assessed. Treating other factors as fixed would have required interpretation of the results to be restricted to the unrealistic and constrained situation in which this experiment was conducted. After all, the 8 tables were chosen by the MP School as its best test of marksmanship. For a crt where task domain is so important, one must be careful not to overgeneralize. However, a random sampling interpretation of the 8 tables could also be made, since they were merely samples from the many distance-position combinations that a policeman may realistically have to shoot at a target.

The comparison of summary statistics shows that the conclusions drawn may at best be incomplete, and at worst erroneous, if only the null hypotheses are tested via F-ratio. The magnitude of effect is a valuable, yet commonly overlooked and ignored summary statistic.

Practical implications. This experimental investigation of a hands-on criterion-referenced test has demonstrated that: (1) replication is a viable way to assess Type I and Type II error rates; (2) experts' judgements about the relative costs of the two error types can be easily assessed, and will likely show that false positives are more serious than false negatives; (3) the sampling assumptions in ANOVA can markedly influence the levels of statistical significance; (4) from a scientific or hypothesis-testing perspective, F-ratios and subsequent null hypothesis acceptance or rejection may be adequate, but from a practitioner's or decision-maker's perspective, the 'practical' significance of an effect needs to be assessed--which can be facilitated by examining the values for the magnitude of effect index.

References

Dodd, D.H., and Schultz, R.F. Jr. Computational procedures for estimating magnitude of effect for some analysis of variance studies. Psychological Bulletin, 1973, 79, 391-395.

Epstein, K.I. Statistical models for criterion-referenced testing and decision-making. Alexandria, VA: U.S. Army Research Institute, Technical Report no. 417, Oct. 1979.

Steinheiser, F.H., and Epstein, K.I. An experimental investigation of the Military Police Firearms Qualification Course. Alexandria, VA: U.S. Army Research Institute Technical Paper no. 322, Sept., 1978.

Validating the Role Reaction Model for Predicting Superior-Subordinate Relationships*

V. Ann Sweney and Arthur B. Sweney
Center for Human Appraisal
Wichita State University, KS

Abstract

The five instruments developed to measure the Role Reaction Model were administered to 117 basic management dyads to superordinates and their subordinate drawn from Titan and Minuteman combat and maintenance crews. The scores for the crew commander's five superordinate roles and five perceptions of specific subordinates were correlated with the crew member's own subordinate roles and his perceptions of his commander. The highest correlation between role groupings were selected to represent the dominant relationship.

The results with minor exceptions, corresponded to the hypotheses created by the Role Reaction Model, ie: Authoritarian Interative roles matched with Ingratiator Interactive; Authoritarian Counteractive with Rebel Counteractive; Permissive Interactive with Rebel Interactive; Permissive Counteractive with Ingratiator Counteractive; and Equalitarian with Critic.

The search for an accurate model for describing superior-subordinate interaction continues. The initial research by Lewin, Lippett, and White (1939) has been followed by research which has focused upon parts of the people system but not all of it. McGregor (1960), Likert (1967), and Blake and Mouton (1964) have focused on the leader's behavior. Fiedler (1963) and later researchers using his model have studied the interaction between leadership behavior and characteristics of the task. Houses (1971) path goal theory acknowledges subordinate relations but does not classify these into formal theory of superior-subordinate interaction.

Sweney (1970) suggested that the Response to Power Model for identifying roles for both superordinates and subordinates. He suggested that Authoritarians and Rebels wanted power, avoid obligation, and responded extrapunitively to frustration. Permissives and Ingratiators want obligation avoid power, and responded to frustration intropunitively. Equalitarians and Critics shared power and obligation, and responded to problems with a non blaming solution-seeking behavior. To neutralize conflict, it was hypothesized that complimentary roles would be sought, ie; Authoritarians with Ingratiators, Permissives with Rebels, and Equalitarians with Critics. The studies to demonstrate this role pairing met with mixed success (Sweney and Fiechtner, 1976). Since conflict was not always reduced, they (Sweney and Fiechtner, 1976) suggested the Role Reaction Model which acknowledged the counteractive and interactive components to the Authoriatarina, Permissive, Rebel, and Ingratiator roles. These make both conflict and complementation possible. The focus of this study was to validate the interactions posited by this model.

*The data for this paper were gathered under a research grant from the Air Force Office of Scientific Research: AFOSR Grant #2001

PRECEDING PAGE BLANK-NOT FILMED

Method

Superordinates and Subordinates were drawn from Titan and Minuteman units to provide 11 dyads. They were given a battery of five instruments developed to measure the Role Reaction Model. They were given consecutive numbers in order to protect their identity but facilitate matching during the analysis and kept as separate units of 62 and 55 dyads respectively.

The instruments were developed to measure self perception, role reactions and perceptions of role partner for the ten roles posited by the model.

The self perception was measured from two different orientations, the pressure role (what was expected) and the preferred role. Since these were relatively new instruments they had not yet achieved the level of reliability which they would later achieve. For this reason, the measured product moment correlations between roles were corrected for attenuation in the manner suggested by DuBois (1965). Due to the large number of correlations a system was devised to locate the dominate relationship. For each of the five roles of the superordinate or subordinate, the highest correlation for that role was drawn from the set of five potential roles of the role partner; ie, Authoritarian Counteractive with Rebel Counteractive, Rebel Interactive, Critic, Ingratiator Counteracitve, or Ingratiator Interacive. The highest correlation was then recorded as representative of the dominate relationship of each referent role.

Results

As can be seen from Table 1, the dominant correlations did define relationships for the subordinate role pressures in eighteen of the twenty-five opportunities and the non hypotheses could be identified as near misses. Similar hit ratios were found on the other seven tables representing other facets of analysis. (leader's role pressure, leader's role preference, subordinate's role preference, for each segment of the sample). Considering that the hypothesized role would appear in the cell only one out of five based on chance alone, the confirmation was strong enough to require no further statistical proof.

Discussion

Examination of the "misses" suggest some interesting speculations. For example, why would the subordinate who feels he should be a Rebel Counteractive have an Ingratiator Interactive Role preference when the reverse is not true? Some of the slippage may reside in the development of the instruments or there may be some more subtle dynamics not hypothesized in the model. This research and earlier research had indicated that role pressure and role preference do not always coincide and that valuable insights result from treating them differently.

Summary

Five instruments for measuring the Role Reaction Model were given to 234 Air Force Missile Men. The data were analysed by pairs of superordinates and subordinates to yield correlations between all scales. The dominate correlations for each role was tabulated to represent the prevailing relationships. Those found corresponding to the model 70% of the time when only 20% would occur by chance.

ROLE REACTION MODEL

Figure 1

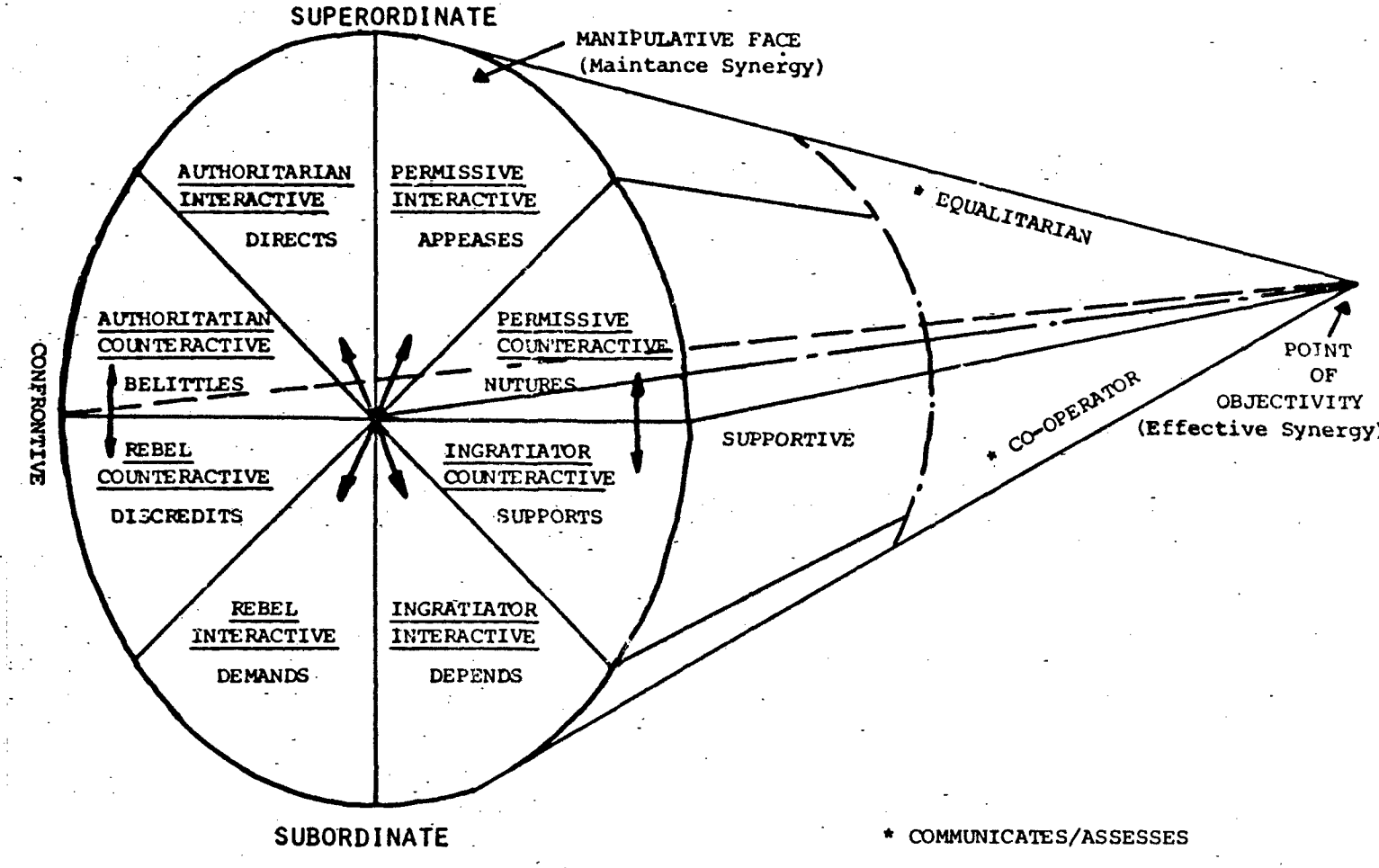


Table 1: Role Relationships Within Titan Dyads
Based upon a Subordinate's Role Pressure (SSRT)
N = 62

Subordinate's Role Pressure (SSRT)	Subordinate's Role Preference (RRT)	Superordinate's Perception of Subordinate (SMR)	Perceived Role of Superordinate (SPR)	Superordinate's Role Pressure (SSRT)	Superordinate's Role Preference (RRT)
REBEL COUNTERACTIVE	N INGRATIATOR INTERACTIVE $r = .31^*$ $r_c = .63^{**}$	H REBEL COUNTERACTIVE $r = .25^*$ $r_c = .32^{**}$	H AUTHORITARIAN COUNTERACTIVE $r = .52^{**}$ $r_c = .65^{**}$	N AUTHORITARIAN INTERACTIVE $r = .31^*$ $r_c = .43^{**}$	H AUTHORITARIAN COUNTERACTIVE $r = .03$ $r_c = .05$
REBEL INTERACTIVE	H REBEL INTERACTIVE $r = .43^{**}$ $r_c = .92^{**}$	H REBEL INTERACTIVE $r = .39^{**}$ $r_c = .62^{**}$	H PERMISSIVE INTERACTIVE $r = .31^{**}$ $r_c = .47^{**}$	H PERMISSIVE INTERACTIVE $r = .25^*$ $r_c = .52^{**}$	H PERMISSIVE INTERACTIVE $r = .26^*$ $r_c = .53^{**}$
CRITIC	H CRITIC $r = .24^*$ $r_c = .41^{**}$	H CRITIC $r = .40^{**}$ $r_c = .57^{**}$	H EQUALITARIAN $r = .51^{**}$ $r_c = .75^{**}$	N PERMISSIVE COUNTERACTIVE $r = .20$ $r_c = .29^*$	N PERMISSIVE COUNTERACTIVE $r = .33^{**}$ $r_c = .78^{**}$
INGRATIATOR COUNTERACTIVE	H INGRATIATOR COUNTERACTIVE $r = .23^*$ $r_c = .50^{**}$	N CRITIC $r = .25^*$ $r_c = .45^{**}$	H PERMISSIVE COUNTERACTIVE $r = .28^*$ $r_c = .51^{**}$	N PERMISSIVE INTERACTIVE $r = .30^*$ $r_c = .55^{**}$	H PERMISSIVE COUNTERACTIVE $r = .16$ $r_c = .48^{**}$
INGRATIATOR INTERACTIVE	H INGRATIATOR INTERACTIVE $r = .22^*$ $r_c = .53^{**}$	H INGRATIATOR INTERACTIVE $r = .27^*$ $r_c = .52^{**}$	N EQUALITARIAN $r = .27^*$ $r_c = .43^{**}$	H AUTHORITARIAN INTERACTIVE $r = .12$ $r_c = .20$	H AUTHORITARIAN INTERACTIVE $r = .25^*$ $r_c = .52^{**}$

H = Hypothesized; N = not hypothesized; * = .05; ** = .01; r = correlation; r_c = corrected correl.

References

Blake, R.R. & Mouton, Jane S. Managerial Grid. Houston: Gulf, 1964

DuBoise, P.H. An Introduction to Psychological Statistics. NY: Harper, 1965

Fielder, F.E. A Contingency Model for the Prediction of Leadership Effectiveness. Technical Report 10, Group Effectiveness Research Laboratory, University of Illinois, Urban, May, 1963

House, R.J., Filley, A.C., & Gujarati, D.N. Leadership Style, Hierarchical Influence and the Satisfaction of Subordinate Role Expectations: A Test of Likert's Influence Propostion, Journal of Applied Psychology. 1971, 55, 422-432

Lewin, K., Lipper, R., & White, R.K. Patterns of Aggressive Behavior in Experimentally Created Social Climates, Journal of Social Psychology. 1939, 10, 271-301

Likert, R. The Human Organization: Its Management and Value. NY: McGraw-Hill. 1967

McGregor, D. The Human Side of Enterprise. NY: McGraw-Hill, 1967

Sweney, A.B. Responses to Power Model Handbook. Wichita: Test Systems, 1971

Sweney, A.B. & Fiechtner, L.A. Factors of Subordinate Superordinate Interaction, Multiveriate Experimental Clinical Research. 1976, 2, 93-100

Sweney, A.B. & Fiechtner, L.A. The Role Reaction Model and Some Confirming Evidence. Technical Report 121, Air Force Office of Scientific Research Center of Human Appraisal, Wichita State University, May, 1976

Time Estimation: Repeated Measures Testing and Drug Effects

Alvah C. Bittner, Jr. and Robert S. Kennedy
Naval Aerospace Medical Research Laboratory Detachment
New Orleans, LA

Michael E. McCauley
Canyon Research Group, Inc.
Westlake Village, CA

Abstract

Using a constant stimuli paradigm, a time discrimination test was administered to seven Navy enlisted men three times in each of four treatment conditions - no drug, amphetamine, hyoscine and amphetamine, and placebo. The twelve test sessions were randomly ordered. The results were analyzed in terms of two Signal Detection Theory (SDT) related metrics, d'-slope and Point of Subjective Equality (PSE), and indicated significant PSE differences between drug treatment conditions ($p < .005$). These effects were superimposed on a strong independent trend toward increasing positive PSE with repeated test administrations ($p < .00001$). The results were discussed in regard to their implications for the simple clock model of drug effects.

The present study involved the reanalysis of previously unreported data on repeated administrations of a time estimation task. The data suggested that an additional component, response bias shift, may contribute to the lack of test stability in time estimation tasks. As originally conceived, the study was intended to investigate the relative effects of two anti-motion sickness drugs on time estimation.

The potential reduction in motion sickness to be derived from drugs must be evaluated relative to their possible undesirable side effects, particularly any decrement in performance. Several types of drugs have been found to be effective against motion sickness, especially some antihistamines and belladonnas (Brand & Perry, 1966; Money, 1970; Wood, Graybiel, & Kennedy, 1966; Wood, Kennedy, Graybiel, Trumbull & Wherry, 1966). Hyoscine (Scopolamine) has been very effective against motion sickness, but side effects of drowsiness, vertigo and dry mouth have been reported when administered orally (Kennedy, Wood, Graybiel, & McDonough, 1966). Recent research has indicated that transdermal administration of hyoscine resulted in a lower frequency of side effects (McCauley, Royal, Shaw & Schmidt, 1979). Combining amphetamine with hyoscine has been reported to enhance effectiveness while reducing side effects (Wood, et al., 1966). Previous research on the effects of amphetamine on performance has shown that overall activity level as well as time judgments change (Goldstone, 1967; Goldstone, Boardman & Lhamon, 1958). In general, overall activity level is increased and subjects tend to overestimate the rate of passage of time. Studies on the effects of hyoscine and barbiturates have reported opposite results, although the data are less consistent (Goldstone & Kirkham, 1968; Kennedy, et al., 1966). In the

PRECEDING PAGE BLANK-NOT FILMED

present study, amphetamine was compared to a combination of hyoscine and amphetamine for effects on time judgment.

Drug induced changes in time judgment may be important for military personnel whose tasks require accurate determination of spatial movement, such as target interception or tracking. It has been suggested that time estimation and the perception of motion and velocity may be related psychological processes (Henderson, 1971; Rachlin, 1966). However, the research literature on time estimation is voluminous (see Zelkind & Sprug, 1971), and includes at least four major methods (production, reproduction, estimation, and comprehension), that often have produced different results. Studies of time perception have used time intervals ranging from less than 50 msec to decades, indicating that a number of different psychological processes have been amalgamated into the construct (McCauley, Kennedy, & Bittner, 1979).

The present study investigated auditory time judgments on the order of four seconds under two drug and two control conditions using a repeated measures test design. The intent of the design was to examine the character of changes in time judgment which could be induced by anti-motion sickness drugs.

Method

Subjects

The subjects were seven young (19-21) enlisted men who volunteered for duty as research subjects at the Naval Aerospace Medical Institute, Pensacola, Florida. These men received extra pay for this duty. All were healthy and motivated. They were treated in accordance with provisions and guidelines concerning the protection of human subjects.

Apparatus

The apparatus consisted of a sound generator, three timers (accurate to one-hundredth of a second), seven head sets, and seven sets of yes/no response keys. The sound generator produced all tones at approximately 900 Hz at 60 dB.

Procedure

Drug Conditions. Two drug conditions and two control conditions were used: (a) amphetamine (10 mg); (b) a combination (H&A) amphetamine (10 mg) and hyoscine (0.5 mg); (c) placebo; and (d) no-drug. One hour before testing, the drugs and placebos were administered in opaque white capsules by a physician. A double-blind procedure was followed, neither the subjects nor the experimenter knew which drug condition had been assigned.

Time Estimation Test. The method of constant stimuli was used for the time estimation test (Woodworth & Schlosberg, 1954; Kling & Riggs, 1971). On each trial a standard (St) tone of 4.0 seconds duration was followed, after a 2.0 second delay, by a comparison (Co) tone. Six Co tones were selected in a pilot study. Their duration ranged from 3.4 to 4.4 seconds, in 0.2 second steps. Subjects were instructed to respond manually with a "Yes" if they judged the Co tone to be longer than the St

tone. Responses of "doubtful" or "undecided" were not allowed. On the rare occasions when they did occur, the trial was repeated.

A repeated measures design was used. All seven subjects were tested simultaneously, but no knowledge of results was given. There were 12 test sessions, three in each of the four drug conditions. The 12 sessions were ordered randomly, and there were approximately 48 hours between sessions. During a session, 15 determinations were obtained on each of the 6 Co tones (i.e., 90 trials per session for each subject). The total data pool consisted of 7,560 comparison trials: 15 trials by 6 Co tones by 7 subjects by 12 sessions (3 in each drug condition).

Results

Two group statistics, Point-of-Subjective-Equality (PSE) and d'-slope were calculated for each of the twelve sessions by least-squares analysis of the average normal (Z) transformation of the response proportions (Woodworth & Schlosberg, 1959; Engen, in Kling & Riggs, 1971, pp. 27-29). As illustrated in Figure 1, PSE is the point where the fitted line crosses the horizontal axis and indicates bias in response when not equal to the standard level (4.0). As also can be seen, the slope of the fitted line, "d'-slope", is a measure of the sensitivity with changes in Co stimuli.* It equals the slope of the line relating changes in Signal Detection Theory (SDT) sensitivity (d') to Co values (and derives its name) because of the linear relationship of Z to d' for any Co, i.e. d' = Z(Co)-Z(St) where Z(Co) and Z(St) are Z-transforms for the response probabilities for Co and St. PSE and d'-slope, it is pertinent to note, characterize the data shown in Figure 1 because the relationship is essentially linear ($r = .99$). The linear characterization held for the data from all 12 sessions with $r > .96$ for all but Day 1 in which $r = .88$. Figure 2 shows the d'-slope and PSE values over the 12 sessions.

Examining Figure 2, the results for PSE appear to present a clear pattern of a linear rise in PSE over sessions with simple drug effects added to the linear rise. This view is supported by the overall fit of a model expressing each score as a linear function of sessions with simple additive drug effects ($R^2(4,7) = .958$; $p < 10^{-4}$). Additional analyses for the sessions and drug effects, mutually partialling out the other, respectively yielded $R^2(1,7) = .498$ ($p < 10^{-3}$) and $R^2(3,7) = .227$ ($p < .005$) indicating significant independent contributions. Figure 3 shows the simple drug effects with no drug and amphetamine conditions being approximately equal and the placebo and H+A conditions having positive effects of 0.11 and 0.14 seconds, respectively. These effects can be contrasted with the change in PSE due to sessions of 0.319 seconds over 12 days (cf. Figure 2).

The d'-slope results, seen in Figure 2, appear to follow no pattern. Fitting of the scores with a model assuming a linear session and simple additive drug effects supported this view with $R^2(4,7) = 0.097$ ($p > .55$). An additional attempt at fitting the data with a simple session effect

*d'-slope can be seen to equal the reciprocal of the traditional measure, σ which is defined as the change in Co with changes in Z from 0 to 1.

yielded similar results with r = .353 (p > .26). Hence, d'-slope appears to be relatively constant while PSE varies systematically.

Discussion

The present investigation examined the effects of drugs and sessions on a discrimination task where standard (St) and comparative (Co) were external to the subject. In addition, a statistical methodology was utilized which yielded results related to Signal Detection Theory (SDT) sensitivity (d') and bias (β) metrics: d'-slope and PSE. The paradigm employed differed from conventional methods for studying drug-time effects which do not use either a St-Co discrimination task or a SDT based analysis.

The results revealed, in addition to lack of changes in d'-slope, pronounced changes in PSE, with sessions and drugs. In particular, the effect of sessions was to move the PSE, initially biased to less than the standard, toward and slightly past it with a change of 0.319 seconds. This result parallels that reported by Needham (1934) who studied intensity judgments of tones over sessions. The effect of drugs was to differentially bias the PSE with amphetamine slightly negatively biased relative to the no-drug condition, and placebo and H&A conditions positively biased 0.11 seconds and 0.14 seconds. Similar to the results reported in previous studies (e.g., Goldstone, 1967; Goldstone & Kirkham, 1968) who have also noted "sedative" placebo effects and opposite dextro-amphetamine effects, the present results were surprising in terms of conventional biological clock theories.

Previous findings of drug effects, using methods investigating subjects' internal time standards, have typically attributed drug effects to changes in the subjects' biological clocks. Tranqualizers and sedatives have typically been hypothesized to slow down subjects' clocks, and psychic energizers (e.g., amphetamine) to speed them up (e.g., Goldstone, 1967). The non-significant d'-slope results are not supportive of the clock hypothesis, however, as changes in clock-speeds would be expected to affect the precision of comparison judgments (faster clocks yielding more precision than slower). Moreover, the finding of drug effects on PSE were counter to clock model predictions as changes in clock speed would not be expected to affect response bias. The current results, therefore, bring into question the simple interpretations of previously reported results. Further, the paradigm employed in this investigation appears as a useful adjunct to current time study methods because of the implications of its application to the attribution of drug effects.

References

Brand, J. J. & Perry, W. L. M. Drugs used in moton sickness. Pharmacological Review, 1966, 18, 895-924.

Goldstone, S. The human clock: A framework for the study of healthy and deviant time perception. Annals of the N.Y. Academy of Science, 1967, 138, 767-783.

Goldstone, S., Boardman, W. K., & Lhamon, W. J. Effect of quinalbarbitone, dextroamphetamine and placebo on apparent time. British Journal of Psychology, 1958, 29.

Goldstone, S., & Kirkham, J. E. The effects of secobarbital and dextroamphetamine upon time judgment: Intersensory factors. Psychopharmacologia, 1968, 13, 65-73.

Henderson, D. C. The relationships among time, distance, and intensity as determinants of motion discrimination. Perception and Psychophysics, 1971, 10, 313-320.

Kennedy, R. S., Wood, C. D., Graybiel, A., & McDonough, R. B. Side effects of some anti-motion sickness drugs as measured by psychomotor test and questionnaires. Aerospace Medicine, 1966, 37, 4.

Kling, J. W. & Riggs, L. A. (Eds.). Woodworth and Schlosberg's Experimental Psychology. New York: Holt, Rinehart, & Winston, 1971.

McCauley, M. E., Royal, J. W., Shaw, J. E., & Schmidt, L. G. Effect of transdermally administered scopolamine in preventing motion sickness. Aviation, Space and Environmental Medicine, 1979, 50, 1108.

McCauley, M. E., Kennedy, R. S. & Bittner, A. C., Jr. Development of Performance Evaluation Tests for Environmental Research (PETER): Time estimation test. Proceedings of the Human Factors Society 23rd Annual Meeting, 1979, 513-517.

Money, K. E. Motion sickness. Physiological Review, 1970, 50, 1-39.

Needham, J. G. The time-error as a function of continued experimentation. American Journal of Psychology, 1934, 46, 558-567.

Rachlin, H. C. Scaling subjective velocity, distance, and duration. Perception and Psychophysics, 1966, 1, 77-82.

Wood, C. D., Graybiel, A., & Kennedy, R. S. A comparison of effectiveness of some anti-motion sickness drugs using recommended and larger than recommended doses as tested in the slow rotation room. Aerospace Medicine, 1966, 37, 259-262.

Wood, C. D., Graybiel, A., Trumbull, R. & Wherry, R. J. Journal of the American Medical Association, 1966, 198, 1155-1158.

Woodworth, R. S., & Schlosberg, H. Experimental Psychology (Rev. ed.). New York: Henry Holt & Co., 1954.

Zelkind, I. & Sprug, J. Time research: 1,172 studies. Metuchen, N. J.: Scarecrow Press, 1974.

Footnotes

The opinions are those of the authors and do not necessarily reflect those of the Department of the Navy.

This research was performed under Navy Work Unit No. MF58.524.002-5027.

Figures

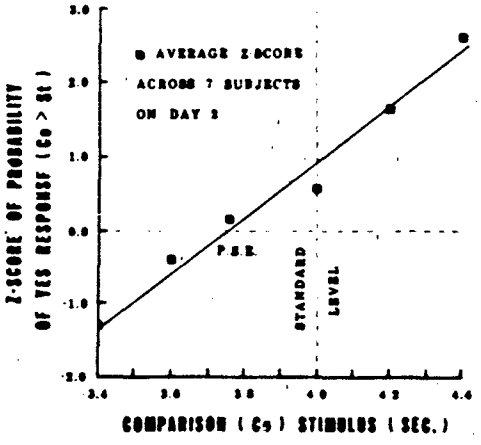


Figure 1. Illustration of P.S.E. and d'-slope.

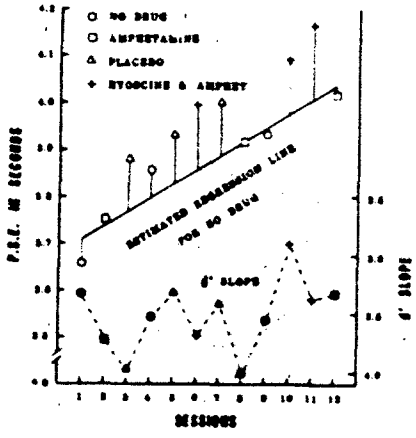


Figure 2. Effects of sessions and drugs on P.S.E. and d'-slope.

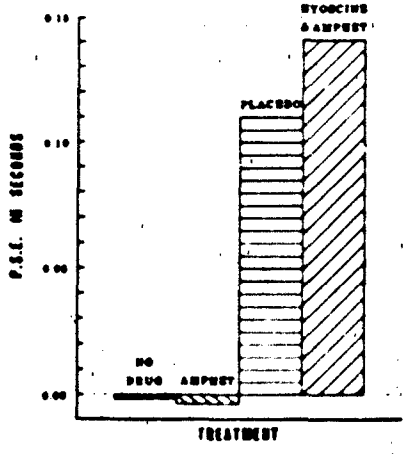


Figure 3. Drug effects on P.S.E.

Performance Evaluation Tests for Environmental Research (PETER): Code Substitution Test

Ross L. Pepper
Naval Ocean Systems Center
Kailua, Hawaii

Robert S. Kennedy and Alvah C. Bittner, Jr.
Naval Aerospace Medical Research Laboratory Detachment
New Orleans, Louisiana 70189

Steven F. Wiker
Department of Industrial Engineering
University of Michigan, Ann Arbor 48104

Abstract

A Code Substitution Test was considered for inclusion in the Performance Evaluation Tests for Environmental Research (PETER) battery. The effects of repeated testing on code substitution performance was studied to determine reliability and stability of task performance. A single two minute testing trial per day was administered to a group of 19 subjects for 15 consecutive weekdays. In a second experiment, a four minute per day test was administered to 12 of the 19 original subjects for an additional 15 consecutive weekdays. Descriptive statistics are reported. Comparisons are made between these laboratory data and performances assessed at sea with repeated administration occurring within each day. The need for knowledge about task stability over repeated performance testing in exotic environments is discussed. The Code Substitution Test is recommended for inclusion in the PETER battery.

A research program is underway to evaluate tests of mental work for future use in studying adverse environments (Kennedy & Bittner, 1977). Each test is examined for stability as it is performed over periods of extended practice (15 days). Tests found to be suitably stable and to possess other characteristics (Kennedy, Bittner & Harbeson, 1980) are made part of a battery of Performance Evaluation Tests for Environmental Research (PETER). The present study reports the findings for a form of the Code Substitution (or Digit-Symbol) Test.

Otis is generally given credit for the initial development of a Digit-Symbol Test, and with Terman, the evolution of group intelligence testing around World War I (Wechsler, 1958). Wechsler (1958) included the Digit-Symbol Test in the original Wechsler-Bellevue (W-B) IQ Test. He felt this inclusion was required because it was one of the oldest and best established of all psychological tests. He felt that the Digit-Symbol Test measured both speed and power, and that both should be given weight in the evaluation of intelligence. He reported high correlations

The opinions are those of the authors and do not necessarily reflect those of the Department of the Navy.

This research was performed under Navy Work Unit No. MF58.524.002-5027.

between Digit-Symbol Test scores and total IQ scores (r = .673 for ages 20-34; r = .697 for ages 35-49 (see Wechsler, 1939 p. 136)). In describing the standardization of his test, Wechsler reported split-half coefficients ranging from r = .83 to r = .90 after correction for attenuation. However, it should be noted that his standardization procedure was not a conclusive demonstration of either reliability, stability, or validity. Correlations within and between the verbal and performance sub-tests indicated the measurement of common variation which could be either a common cluster of factors, correlated errors of measurement within days, or both. Hence, the consistency of the Digit-Symbol Test is not clear.

In addressing this issue, Derner, Aborn and Cantor (1950) rightly pointed out that the method of choice for determining the reliability of a measuring instrument is a test-retest technique. They then conducted a test-retest study to assess changes over 6 months, 4 weeks, and 1 week using normal adults (n=158). In all sub-tests, including Digit-Symbol, a learning effect was apparent. The overall WAIS reliability coefficients across test-retest intervals varied from r=.83 to r=.88 for the performance scale and Digit-Symbol was r=.80. This was the first substantial evidence that the Code Substitution Test has sufficient reliability to potentially reflect changes with environmental manipulations. It is noteworthy that except for the schizophrenic population employed by Ragin, all adult reliabilities on the Digit-Symbol test surveyed by Derner, et al. (1950) exceeded the mid .70's . Hence, the body of literature suggests that the Digit-Symbol test has adequate simple test-retest reliability.

The stability of the Digit-Symbol test alone across extensive repeated testing or practice, has not been sufficiently established in previous research. The most relevant study was by Woodrow (1937) who compared the performance of high and low initial score performers on a variety of tests, including a Code Substitution Test. Testing was conducted daily for a 10 minute period for 39 days for one group (n = 56) and for 66 days for a second group (n = 82). The initial-final reliability coefficients for code substitution were r = .57 for the former, r = .59 for the latter. The ratio of initial and final group standard deviations were 1.57 and 1.64 respectively for the two groups, indicating that between subject differences increased slightly with practice, a finding that has been obtained elsewhere (Harbeson, Kennedy, & Bittner, 1979). The extent to which performance on a variety of tasks confounds findings is not known. Therefore, the primary purpose of the present effort was to study code substitution in the laboratory under baseline conditions over extended practice. A secondary purpose was to report the sensitivity of this test in a field study.

Experiment 1

Method

Subjects. Navy enlisted men (n=19) age 19-24 comprised the experimental group. These men were recruited, evaluated and employed in accordance with procedures described elsewhere (Thomas, Majewski, Ewing &

Pristo (1978) has shown lower test-retest reliabilities ($\bar{r}$ = .20) in 40 children (IQ range 52-145) than expected.

Gilbert, 1978). These procedures meet or exceed prevailing national and international guidelines concerning human use in research. The subjects received extra compensation for volunteering and appeared motivated to perform. They were representative of the Navy population in size and intelligence but physically and mentally screened for hazardous duty environment research. They were under continuous medical supervision.

Apparatus. The Code Substitution test forms were derived after the concepts of Otis, where each day nine letters were randomly assigned a digit from one to nine. Fifteen alternate forms were computer generated following a general Monte Carlo algorithm: (a) the digit letter relationships were changed daily; (b) each letter appeared 10-15 times in a daily list of 135 items; and (c) each letter was nonrepeating. Figure 1 shows a layout of a sample test form.

Procedure. The subject's task was to follow the letter/number correspondence for a given day in assigning the appropriate letter below each number. Subjects were instructed to proceed rapidly and accurately throughout the list until told to stop. Each session in Experiment 1 lasted two minutes. The subjects were ordinarily tested in a group each workday morning for three weeks. Performance was scored according to number attempted, number correct, and rights minus wrongs. Group means, between subject standard deviations, and cross session reliabilities were calculated for each score. Analysis of variance (ANOVA) was conducted for days and subjects main effects.

Results

Only results for total-correct are reported here as the subjects made very few errors, (1 on the average/per subject/per day) and other scores (e.g. total attempted) were redundant. Figure 2 shows means and standard deviations for total-correct for nineteen subjects over 15 days. Mean performance is seen to improve throughout the study, although the trend becomes less pronounced after Day 8. Similarly, standard deviations increase but are relatively constant after Day 8. Figure 3 shows the cross session reliabilities for selected base days, the source of which is Table 1. Correlation traces (Bittner, 1979) show negative slopes for Base Days 1, 2, and 4. This trend is less evident in traces for Base Days 8, 10, and 12, suggestive of differential stabilization somewhere between Days 4 and 8. Task definition (Jones, 1980), the degree to which a test differentiates reliably between individuals, is greater than $r = .75$ subsequent to Day 8.

Experiment 2

Method

Subjects. Twelve of the 19 original subjects comprized the experimental group. Between the end of Experiment 1, and the beginning of Experiment 2, the other 7 subjects were transferred and were not available for testing.

Apparatus. The test forms were produced in the same way as in Experiment 1, with the exception that each day's test was twice as long (270 vs. 135 items).

Procedure. The procedure was the same as Experiment 1, except that the subjects were given 4 minutes rather than 2 minutes of testing each day. The testing period began 11 weeks after the conclusion of the first experiment, and continued for 15 consecutive workdays.

Results

Experiment 2 was conducted in an attempt at improving the magnitude of the correlation level by doubling testing time. Although only twelve of the original 19 subjects remained available for the retest, their means (Figure 5) were not statistically different from the original group ($p > .5$). The second study also was continued for fifteen days, and the means and standard deviations for these twelve subjects appear in Figure 6. While performance continued to improve over the period of the experiment, the change is slight but significant ($p < .01$). Not unexpectedly, the values are about twice those of the shorter test (cf. Figure 3). Correlations are level for all comparisons indicating task stabilization was manifested on Day 1 of Experiment 2. Task definition is better than with the shorter test b·t slightly less than predicted by a Spearman-Brown adjustment.

Experiment 3

Method

Subjects. Six U. S. Coastguardsmen were selected from the complement of the WPB 95 (White Patrol Boat) employed in this study.

Apparatus and Procedure. Testing materials and procedures were similar to those employed in Experiments 1 and 2 with the following exceptions: Testing was conducted hourly from 0800-1600 for four consecutive days. The testing compartment was located amidships, below decks. The first two days of testing were conducted dockside, with engines running. The second two days of testing occurred while the vessel steamed a double octagonal pattern seven miles southwest of Honolulu in the Molokai Channel, an area acknowledged for its turbulent sea condition. The testing commenced each sea day while the vessel steamed directly into the primary swell. Course changes of 45° were made every half hour throughout the day, creating a systematically changing motion environment. (See Wiker & Pepper, 1978 for greater details of the testing conditions and a description of other task and subject variables assessed during this phase).

Results

Figure 7 shows performance on the Code Substitution Test for the six Coastguardmen exposed to heavy seas in the Molokai channel. The data are plotted as scores per minute for the 16 dockside practice trials versus the 16 at sea data points. For comparability, the data from the first and second laboratory studies (Figures 4 and 5) have been replotted as a function of cumulative practice. Plotted in this way, 15 days of 2 minute laboratory trials can be compared to the first 15 hours of dockside testing.

The fit between the two studies for the first 30 minutes of practice is surprisingly good considering the known differences in the two experiments: (a) design - all performance massed in 4 days versus distributed over two 3-week periods 11 weeks apart; (b) test length - 2 and 4 minute trials were combined in the laboratory study versus two minute trials only in the field study; and (c) subjects - Navy versus Coastguardsmen. Secondly, the fit is also good during the sea trials with the exception of the second hour at sea where the poorest performance of all was obtained. This finding of performance degradation is concordant with the high motion sickness symptoms during this time frame (Wiker, Kennedy,

McCauley & Pepper, 1979). Moreover, because of the stability and differentiation of the laboratory version of the task and the close agreement between the two studies after the at sea decrement, the authors are inclined to consider this a real effect of motion on performance.

Discussion

The PETER Program is underway whereby psychological tests are being examined critically to determine their suitability for use in detecting performance degradation in novel environments (Kennedy & Bittner, 1977). The criteria against which tests are compared focus on stability and sensitivity. Stability is measured by examining the effects of extended practice on means, standard deviations and cross session reliabilities. Means are stable if they are level, asymptotic or exhibit constant slope. Standard deviations may be level or increase slightly with the mean. Cross session reliabilities are considered stable after they cease to change over sessions. The 2 and 4 minute tests qualify for PETER. Means have constant slope after Day 8 of Experiment 1 and standard deviations are also level after that time. The reliabilities are moderate $\bar{r} >$ .75 and stable after Day 8. Experiment 2 showed several things: (a) stability is still present 3 months later; (b) a test twice as long only improves reliability to an average of $\bar{r}$=.80 while effectively doubling mean performance. This Code Substitution Test appears to be an excellent candidate for inclusion in PETER from the laboratory results.

The results of the sea trials in this study provide at the same time vindication and validation of the PETER paradigm. The laboratory task sufficiently differentiates subjects, and is stable, so that slight departures may be ascribed as due to environmental and not artifactual variables. The benefit of being able to compare real world performances at sea with those of a control group in a laboratory is also noteworthy. Both laboratory and environmental results recommended the use of the Code Substitution Test in PETER or other environmental batteries.

References

Bittner, Jr., A. C. Tests of differential stability. Proceedings of the 23rd Annual Meeting of the Human Factors Society, Boston, October, 1979.

Derner, G. F., Aborn, M. and Cantor, A. M. The reliability of the Wechsler-Bellevue Subtest and Scales. Journal of Consulting Psychology, 1950, 14, 172-179.

Harbeson, M. M. Kennedy, R. S. & Bittner, A. C., Jr. A comparison of the Stroop Test to other tasks for studies of environmental stress. Proceedings of the 12th Annual Meeting of the Human Factors Association of Canada, Bracebridge, Ontario, Canada, September, 1979.

Jones, M. B. Stabilization and task definition in a performance test battery. (NAMRL Monograph No. 27). Pensacola, FL: U.S. Naval Aerospace Medical Research Laboratory, 1980.

Kennedy, R. S. & Bittner, A. C., Jr. The development of a Navy Performance Evaluation Test for Environmental Research (PETER). In, Productivity Enhancement: Personnel Performance Assessment in

Navy Systems. Symposium presented at the Naval Personnel Research and Development Center, San Diego, CA, 12-14 October, 1977. (NTIS No. AD-056047).

Kennedy, R. S., Bittner, A. C., Jr., & Harbeson, M. M. An engineering approach to the standardization of Performance Evaluation Tests for Environmental Research (PETER). Proceedings of the 11th Annual Conference of the Environmental Design Research Association, Charleston, S.C., March, 1980.

Pristo, L. J. Comparing WAIS and WISC-R Scores. Psychological Reports, 1978, 42, 515-518.

Thomas, D. J., Majewski, P. L., Ewing, C. L. & Gilbert, N. S. Medical Qualification Procedures for Hazardousduty Aeromedical Research. (Conference Proceedings No. 231, A3, pp. 1-13, 1978) London: AGARD, 1977.

Wechsler, D. Measurement of Adult Intelligence (1st ed.). Baltimore: Williams & Wilkins Co., 1939.

Wechsler, D. The Measurement and Appraisal of Adult Intelligence. Baltimore: The Williams & Wilkins Co., 1958.

Wiker, S. F. & Pepper, R. L. Change in crew performance, physiology and affective state due to motions aboard a small monohull vessel; a preliminary study. Coast Guard Technical Report No. CG-D-75-78, 1978.

Wiker, S. F., Kennedy, R. S., McCauley, M. E., & Pepper, R. L. Susceptibility to seasickness: Influence of hull design and steaming direction. Aviation, Space & Environmental Medicine, 1979, 50, 1046-1051.

Woodrow, H. Factors in improvement with practice. The Journal of Psychology, 1937, 7, 55-70.

Table

Table 1
Experiment 1: Comparison of Reliabilities for Total Correct Over 15 Days (n=18)

Days	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	.60	.50	.33	.66	.70	.67	.56	.65	.70	.57	.63	.70	.65	.53
2		.74	.53	.72	.60	.62	.68	.50	.61	.61	.51	.53	.50	.53
3			.77	.61	.62	.60	.61	.58	.74	.57	.62	.66	.62	.67
4				.73	.51	.70	.70	.68	.82	.67	.55	.70	.54	.66
5					.66	.73	.67	.77	.80	.70	.71	.75	.66	.70
6						.70	.70	.57	.70	.66	.75	.67	.66	.76
7							.50	.65	.56	.66	.73	.56	.72	.65
8								.67	.90	.77	.66	.76	.65	.69
9									.72	.71	.73	.77	.66	.76
10										.66	.76	.66	.66	.76
11											.71	.65	.56	.76
12												.70	.75	.65
13													.77	.67
14														.65

Figures

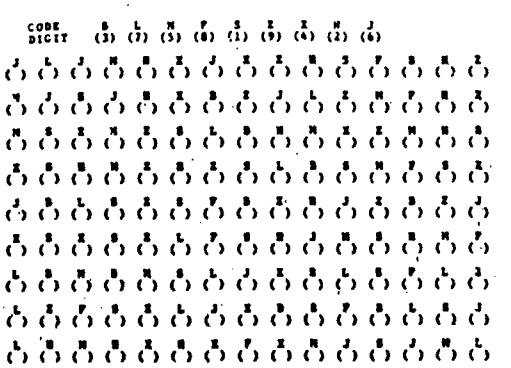


Figure 1. Coding Test Sample Form (Day 1, Experiment 1)

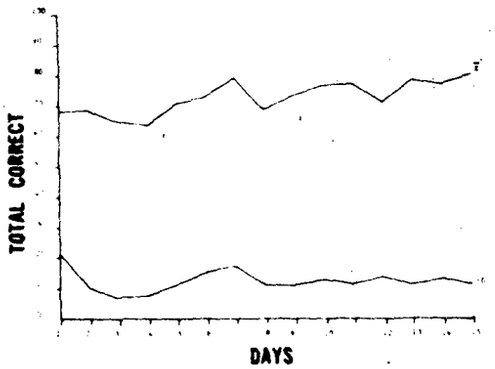


Figure 2. Experiment 1: Means and Standard Deviations for Total Correct Over 15 Days (n=19).

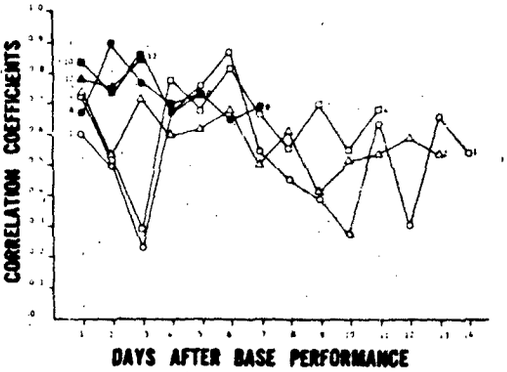


Figure 3. Experiment 1: Correlations Between Selected Base Days (1, 2, 4, 8, 10, 12) and Those Following For Total Correct Over 15 Days (n=19).

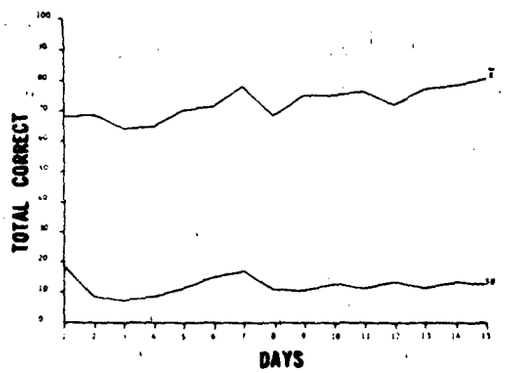


Figure 4. Experiment 1: Means and Standard Deviations for Total Correct Over 15 Days (n=12).

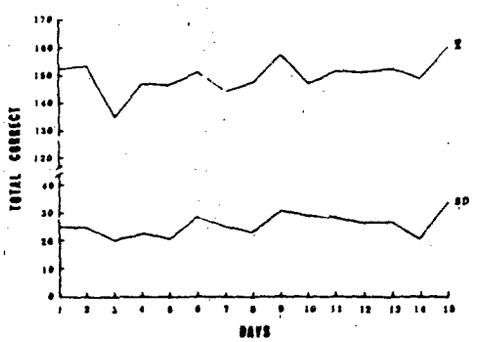


Figure 5. Experiment 2: Means and Standard Deviations for Total Correct Over 15 Days (n=12).

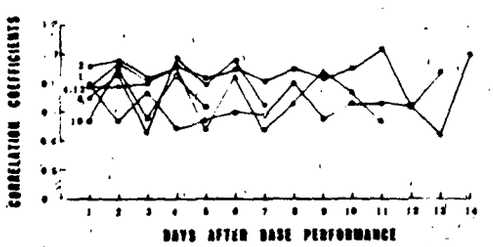


Figure 6. Experiment 2: Correlations Between Selected Base Days (1, 2, 4, 8, 10, 12) and Those Following for Total Correct Over 15 Days (n=12).

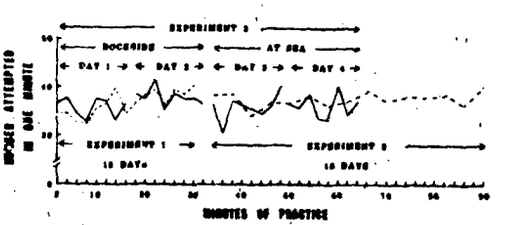


Figure 7. Experiments 1, 2, and 3: Mean Number Attempted in 1 Minute.

Performance Evaluation Tests for Environmental Research (PETER): Interference Susceptibility Test (IST)

Michele Krause and Robert S. Kennedy
Naval Aerospace Medical Research Laboratory Detachment
New Orleans, Louisiana

Abstract

A program designed to develop Performance Evaluation Tests for Environmental Research (PETER) is in progress. Underwood's (1977) Interference Susceptibility Test (IST) was evaluated for inclusion in PETER on the basis of its suitability for repeated administrations. Baseline testing consisted of alternate forms of the IST being administered to 23 subjects for 15 workdays. The results show the mean of the total percent correct score continues to exhibit a slow increase over the entire experiment, with the standard deviation remaining constant subsequent to Day 7. Reliability correlations appear differentially stable after some training ($\bar{r} \approx .75$). The slope score, the traditional measure of IST, is unreliable, although the standard deviations are relatively constant. The total percent correct score is recommended for possible inclusion in PETER.

The Navy is developing Performance Evaluation Tests for Environmental Research (PETER) at its medical laboratory in New Orleans. The goal of the PETER program is to develop a multiple administration test battery which will be effective in detecting performance decrements that are caused by ship motion. Additionally, due to its nature, the test battery is expected to lend itself to the study of other stressors, such as toxic drugs, extreme temperatures and high pressure. The current phase of this project involves repeated testings of cognitive, perceptual and psychomotor tasks. In choosing a task for study, one or more of the following criteria must have been met: (a) performance has been shown to be disrupted in a thermal, inertial or hyperbaric environment, (b) it has been acknowledged to assess cognitive, information-processing, or memory functions, or (c) normal subjects have been distinguished from brain damaged persons (Kennedy & Bittner, 1977). One of the tasks selected for study was Underwood's Interference Susceptibility Test (IST) (Underwood, Boruch & Malmi, 1977). This task was originally designed by Underwood to study the effects of proactive interference. In this original study, 200 college students were tested on 24 separate tasks. Fernandes and Rose (1978) included the test in their studies of an information-processing approach to performance assessment. It is suspected that the more basic memory tasks which have been studied at NAMRLD (e.g., recall and recognition tasks) do not distinguish memory capacities in the same way as IST does. The Interference Susceptibility Test required associations to be formed, dismissed, and then new, conflicting assocations formed. During

---

The opinions are those of the authors and do not necessarily reflect those of the Department of the Navy.

This research was performed under Navy Work Unit No. MF58.524.002-5027.
The authors are indebted to Andrew Rose for providing stimulus material.

PRECEDING PAGE BLANK-NOT FILMED

exposure to persons suffering from motion sickness, one of the authors found that "confusion" was reported as a frequent mental symptom. It is possible that IST is sensitive enough to measure a component of "confusion".

The purpose of the present study is to determine whether IST is suitable for use in environmental research. From our point of view, a task is considered suitable if it has task definition (i.e. differentiates between subjects) and is stable. In accordance with Jones (1979), stability exists when: (a) the daily group means asymptote or evidence a slight, constant slope, (b) day-to-day variance is constant, and (c) relative performance standings between subjects are constant from day to day. A recommendation of whether to include this test in subsequent PETER studies is based on these criteria. Reviews which describe this program in detail, as well as describe the results of previous tasks that have been administered, are available (Harbeson, Kennedy & Bittner, 1979; Kennedy & Bittner, 1977; Kennedy, Bittner, & Harbeson, 1980).

Method

Subjects

The subjects were a group of 23 volunteer enlisted Navy men, ages 19 to 24. To qualify for this medical research program, they had to be within the norms for Navy enlisted personnel in physical health, mental health and intelligence. All subjects were recruited, evaluated and employed in accordance with procedures specified in Secretary of the Navy Instruction 3900.39 and Bureau of Medicine and Surgery Instruction 3900.6. These instructions are based upon voluntary consent, and meet the provisions of prevailing national and international guidelines. A description of the subject selection procedure is given by Thomas, Majewski, Ewing and Gilbert (1978).

Task description

Stimulus material for each session was comprised of lists of trigram-digit pairs (e.g. NOB-2). A list was made up of five trigrams paired with digits from 1 to 5. During each session, three sets, each containing four lists, were administered. Across the four lists of each set, the same trigrams were paired with digits from 1 to 5, forming different combinations in each list. Stimulus material was provided by Rose. An example of stimulus material for one set is found in Table 1.

Apparatus and procedure.

Subjects were shown each of five trigram-digit pairs by means of a single slide, presented on a Kodak Ektagraph 450 AudioViewer®. The rate of presentation was one slide every 3 seconds. A cueing slide appeared at the end of the list and at the beginning of the recall list. Each trigram was then shown by itself (in an order different from the paired presentation) for 4 seconds, and subjects recorded the number with which they thought each trigram had been paired. Subjects were tested in groups of four, at 8:00 in the morning, for 15 consecutive workdays.

Results

Two measures were taken across sets for four lists: (a) slope of lists and (b) percent correct for each list. In addition, mean percent correct was obtained for each of three sets (summed over lists) and an aggregate mean (over sets and lists) was obtained in order to compare results with Underwood, et al. (1977).

Figure 1 shows the mean percent correct responses across sets for the four lists. As expected, performance declines with each successive list that is presented. The impression of a learning curve over days is observable across each list. The greatest improvement is seen in List 1 (33%). The reason for the anomalous scores on Day 6 is obscure. Standard deviations, as seen in Figure 2, are level and unremarkable.

Percent correct performance for each of the three sets (summed over lists) showed that subjects exhibit a slight advantage for later sets (not shown), although the differences are negligible. Mean performance for the three sets, across lists progresses from 50.1 on Day 1 to 71.8 on Day 15. The average percent correct in both this study and the Fernandes & Rose (1978) study was 65%. Underwood, et al. (1977) obtained an 85 percent correct average when this test was interdigitated with 23 other memory tests.

When Underwood et al. (1977) correlated total correct responses for Sets 1, 3, 5 with those same scores from Sets 2, 4, 6, they obtained a value of $r = .81$. This correlation between successive sets (i.e. split half) in Underwood's study is compared to a correlation of $r = .74$ between successive days (i.e. test-retest) in the present research, wherein the number of observations are the same for both calculations. There is no evidence that the reliabilities of the present data are different from those of Underwood et al. (1977) ($z = .72$, $p > .40$).

Tables 2 and 3 show reliabilities within Lists 2 and 4. Because Lists 1 and 3 revealed comparable results, they are not shown. These correlations reveal that average percent correct performance appears to stabilize around Day 8. This result is, perhaps, more clearly illustrated when Table 2 is graphed as in Figure 3. This figure presents correlations of percent correct performance for selected testing days in a left-justified manner, enabling examination of all subsequent testing days. Although a progression towards stabilization occurs, the task definition remains too low to be satisfactory (Jones, 1979).

Figure 4 shows the means and standard deviations for the slope scores over lists. Mean slopes are variable and show no systematic trend. The standard deviations are equal to the means suggesting substantial differences between subjects. Table 4 shows slope reliabilities. Composite reliability for this score is essentially zero ($r = .04$).

Discussion

Percent correct scores for the individual lists provide evidence for stabilization within the second week of testing, but with task definition at too low a level to be considered useful. When the percent correct scores are summed over lists and sets task definition improves ($\bar{r} = .71$),

and reliabilities after Day 8 appear stable. This aggregate score is the one favored by Underwood, et al. (1977), who found it to be correlated with the slope measure. While less defensible as a measure of interference susceptibility, the percent correct score over lists and sets meets the minimum requirements for suitability for PETER and will be employed in subsequent analyses at this laboratory. It should be noted that the test in its present form, requires ten minutes to complete and yields a composite reliability in List 2 (as an example) of r = .53 Using the Spearman-Brown adjustment formula (Allen & Yen, 1979), reliability raises to r = .69 if the testing length is doubled. The total aggregate score improved from r = .71 to r = .83.

The chief finding in this experiment is that the slope score, theoretically the most meaningful measure of the interference factor, is unreliable (r = .04). This poor reliability over sessions is not due to insufficient variance between subjects and it occurred despite the fact that the slope means and standard deviations are stable. Fernandes and Rose (1978) also obtained low reliability for the slope measure (r = .05). It is probable that the same cautions which are associated with difference scores (Cronbach & Furby, 1970) may apply to slopes. Those authors suggest, as an alternative, analyzing the most complex condition with the simpliest condition as a covariate (in this case, List 4 with List 1 as a covariate). This analysis will be performed on the IST data at a later date.

In conclusion, IST as analyzed up to this point, is not an ideal candidate for inclusion in future PETER studies. It is recognized though, that with some modifications to the adminstration procedure, this test may reveal a unique factor of memory that would be useful to include in the final PETER battery. It may prove to be necessary, when studying other environmental stressors, (specifically impact acceleration) to place heavier emphasis on memory tasks because of the close connection between memory and other human systems and functions.

References

Allen, M. J. & Yen, W. M. Introduction to Measurement Theory. Belmont, California: Wadsworth, Inc., 1979.

Cronbach, L. J. & Furby, L. How should we measure "change" - or should we? Psychological Bulletin, 1970, 74, 68-80.

Fernandes, K. & Rose, A. M. An Information Processing Approach to Performance Assessment: II. An Investigation of Encoding and Retrieval Processes in Memory. (Tech. Report IAR 58500-11/78 TR). Washington, D.C.: American Institutes for Research, November, 1978.

Harbeson, M. M., Kennedy, R. S., & Bittner, Jr., A. C. A comparison of the Stroop Test to other tasks for studies of environmental stress. Proceedings of the 12th Annual Meeting of the Human Factors Association of Canada, Bracebridge, Ontario, Canada, 6-8 September, 1979.

Jones, M. B. Stabilization and Task Definition in a Performance Test Battery. (NAMRL Monograph No. 27). Pensacola, FL: U. S. Naval Aerospace Medical Research Laboratory, 1980.

Kennedy, R. S. & Bittner, A. C., Jr. The development of a Navy Performance Evaluation Test for Environmental Research (PETER). In, Productivity Enhancement: Personnel Performance Assessment in Navy Systems. Symposium presented at the Naval Personnel Research and Development Center, San Diego, CA, 12-14 October 1977. (NTIS No. AD 056047).

Kennedy, R. S. & Bittner, Jr., A. C. Progress in the analysis of Performance Evaluation Tests for Environmental Research (PETER). Proceedings of the 22nd Annual Meeting of the Human Factors Society, Detroit, Michigan, October, 1978. (NTIS No. AD A060676)

Kennedy, R. S., Bittner, Jr., A. C. & Harbeson, M. M. An engineering approach to the standardization of Performance Evaluation Tests for Environmental Research (PETER). Proceedings of the 11th Annual Conference of the Environmental Design Research Association, Charleston, S. C., March, 1980.

Thomas, D. J., Majewski, P. L., Ewing, C. L. & Gilbert, M. S. Medical Qualification Procedures for Hazardous-Duty Aeromedical Research. (Conference Proceedings No. 231, A3, pp. 1-13, 1978) London: AGARD, 1977.

Underwood, B. J., Boraci, R. F. & Malmi, R. A. The composition of episodic memory. (ONR Contract No. N00014-76-C-0270) Evanston, Illinois: Northwestern University, May 1977. (NTIS No. AD A040696)

Tables

Table 1
Stimulus Presentation

	List shown to Ss	Probe shown to Ss	Correct response
LIST 1	DOG - 5	WIN	1
	MOB - 2	PEG	4
	WIN - 1	DOG	6
	PEG - 6	MOB	2
	MEN - 3	MEN	3
LIST 2	WIN - 5	PEG	2
	PEG - 2	MOB	3
	DOG - 4	MEN	1
	MEN -	DOG	4
	MOB - 3	WIN	5
LIST 3	MEN - 2	MEN	2
	MOB - 5	MOB	5
	DOG - 1	DOG	1
	WIN - 6	WIN	6
	PEG - 3	PEG	3
LIST 4	DOG - 3	MEN	5
	MEN - 5	WIN	2
	MOB - 4	PEG	1
	PEG - 1	MOB	4
	WIN - 2	DOG	3

Table 2
Mean Percent Correct
Reliabilities for List 2

Days	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	.32	.06	.42	.01	.07	.43	-.06	.15	.27	.32	.25	.12	-.05	.24
2		.31	.46	.46	.33	.44	.30	.38	.41	.06	.24	.19	.22	.30
3			.62	.28	.00	.53	.58	.27	.68	.57	.50	.72	.70	.64
4				.55	.34	.70	.61	.69	.70	.65	.54	.61	.63	.56
5					.24	.17	.12	.46	.35	.23	.26	.45	.32	.30
6						.55	.24	.51	.41	.10	.28	.28	.05	.24
7							.46	.52	.60	.45	.67	.52	.51	.51
8								.55	.70	.53	.65	.61	.53	.67
9									.70	.64	.51	.51	.46	.32
10										.61	.68	.66	.58	.62
11											.64	.68	.67	.75
12												.53	.55	.66
13													.77	.70
14														.72

Table 3
Mean Percent Correct
Reliabilities for List 4

Days	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	.32	.59	.40	.38	-.16	.34	.07	.60	.33	.32	.58	.39	.48	.44
2		.05	.10	-.07	.18	.43	.08	.15	.52	.42	.35	.16	.14	.20
3			.34	.18	-.43	.14	-.05	.16	.05	-.06	.34	.06	.06	.10
4				.33	-.01	.40	.23	.47	.25	.30	.15	.42	.51	.50
5					.23	.40	.48	.26	.48	.35	.24	.41	.55	.50
6						.09	-.00	.15	.06	.22	-.15	.16	.10	.27
7							.55	.35	.50	.81	.47	.42	.52	.46
8								.18	.34	.44	.09	.37	.55	.02
9									.40	.34	.45	.50	.64	.37
10										.46	.41	.60	.53	.49
11											.40	.55	.54	.46
12												.40	.41	.41
13													.73	.59
14														.44

Table 4
Reliabilities of Mean
Slope of Lists Across Sets

Days	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	-.21	.15	-.14	.06	.03	.06	-.17	.16	.06	-.00	.00	.16	-.15	-.06
2		.10	-.02	-.13	-.05	.32	-.00	-.05	.10	.29	.09	.06	-.05	.02
3			.07	.16	-.10	-.05	.30	-.01	.32	-.31	.16	.01	.21	-.28
4				-.19	.20	.30	-.15	-.11	-.29	.31	-.06	.25	.19	.03
5					-.31	-.13	.36	.06	.62	.32	.40	.57	.56	.06
6						.32	-.12	.23	-.29	-.00	-.16	.07	-.09	.04
7							-.05	.20	-.08	.56	.40	.38	.03	.18
8								-.00	.26	.09	-.18	-.15	.26	-.15
9									.16	.08	.49	.29	-.01	.00
10										.01	.35	.17	.14	.33
11											.45	.62	-.11	.29
12												.60	.29	.14
13													.39	.07
14														.03

Figures

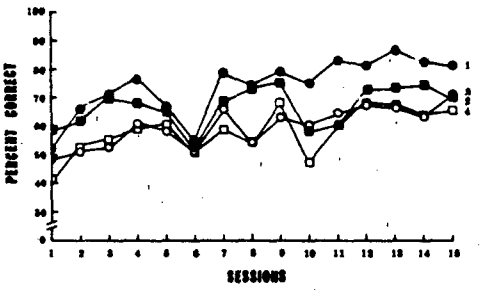


Figure 1. Mean Percent Correct Across Sets for Lists 1, 2, 3, & 4 Over 15 Days.

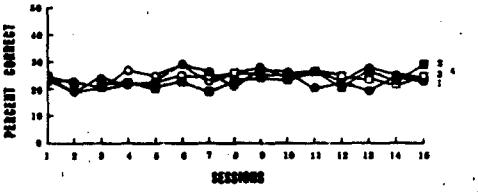


Figure 2. Standard Deviations of Percent Correct Across Sets for Lists 1, 2, 3 & 4 Over 15 Days.

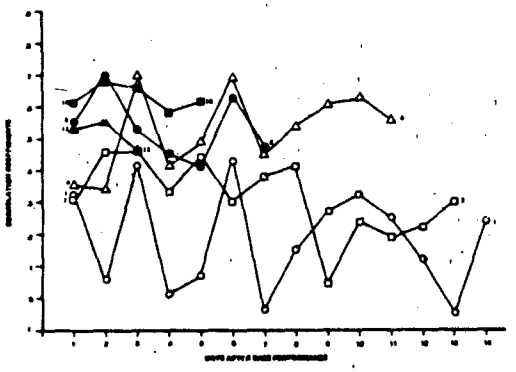


Figure 3. Reliabilities for Percent Correct Across Sets for List 2 for Selected Base Days 1, 2, 3, 4, 8, 10, 12 and Those Following over 15 Days.

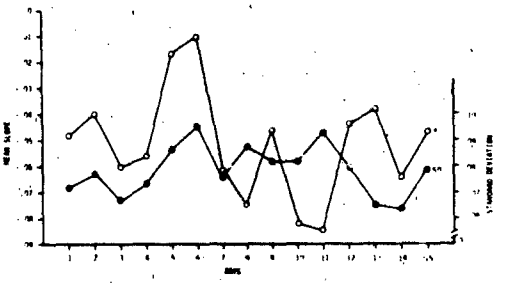


Figure 4. Means and Standard Deviations for the Mean Slope of Lists Across Sets Over 15 Days.

Video Games and Convergence or Divergence with Practice

Marshall B. Jones
The Pennsylvania State University
Hershey, PA 17033

CDR Robert S. Kennedy, MSC, USN

and

Alvah C. Bittner, Jr.
Naval Aerospace Medical Research Laboratory Detachment

Abstract

Video Games

In 1972 a coin-operated video game called Pong and manufactured by Atari, Inc., a company founded that same year, appeared on the electronic-games market. In less than a year Atari sold 6,000 games at more than $1,000 apiece. Midway Manufacturing Co., which Atari licensed to produce a version of Pong, sold 9,000 of the table-tennis type games in less than six months.

Also in 1972, Magnavox marketed a video game called Odyssey that could be played on home TV sets. The Odyssey set included a control unit, which attached to a home TV set and permitted one to play 12 different games by inserting a "game card" into the control unit. The original Odyssey was not, however, a programmable video game. All 12 games were resident in the control unit and were not, in fact, very different; the "game card" set appropriate lines, bars, and cursors. Then in 1975 Atari entered the home video market with a version of Pong that offered several new advances: electronically generated on-screen courts, sound effects for every hit, miss, and ricochet, and automatic on-screen digital scoring. By the end of 1976 twenty different companies, including Coleco, First Dimension, National Semiconductor, Phoenix, Unisonic, and Universal Research were producing video games for home use.

About this time, that is, late in 1976, Fairchild Camera and Instrument entered the field with the first fully programmable video system. The system was programmed by inserting an electronic cartridge into the game console. The benefit was that one could play as many different games as the company provided cartridges -- in fact, more, because most cartridges contained several games. Different games within the same cartridge were selected by punching in a number on the control console.

In 1977 and 1978 programmable video games for home use proliferated on all sides. Companies already in the field, like Atari and Magnavox, came out with programmable video systems; and new companies entered the field, for example, RCA and Bally, the pinball-machine company. In 1978 American shoppers spent more than 200 million dollars on programmable home video games and everything pointed toward an even larger market in the future.

Video games as psychological tests

The potential of programmable video games for psychological testing is large. First, the new games involve skills and lots of them. Video games are tasks and playing them repeatedly constitutes so many trials of practice. The more a person plays the better he or she becomes, especially in the beginning; after extended practice, the gains from playing yet another game are small or non-existent. Most of the games, moreover, have a high ceiling, so high that few people come close to reaching it. Second, the new games are wonderfully self-motivating. A case can be made that for research purposes solid motivation is not all to the good. Insufficient motivation, boredom, or wavering attention may be precisely what the investigator wishes to study, and in such a case video games would not be the tasks of choice. More often, however, we are interested in skill acquisition, learning or forgetting, as distinct from performance; and where we are, insufficient or wavering motivation is quite simply a source of error. Third and last, most video games are highly speeded. In fact, this feature of the games may account for much of their appeal. In considerable measure the games are enjoyable because they operate at more or less the same speeds as we do, that is, as our brains do. Their being so fast, however, may permit them to tap aspects of human functioning that escaped us as long as we were dealing with essentially mechanical tasks (pursuit rotor, two-hand or complex coordination).

Programmable video games are equally attractive at a pragmatic level, especially for performance testing. Literally dozens of games and, in principle, hundreds or even thousands can be played with identically the same equipment; one need only insert another cartridge. Television sets are light, easily transported, and occupy little space. Furthermore, if they break down, they are easily replaced. The game console and associated cartridges are robust. The only parts of game equipment that show any appreciable tendency to break down are the joysticks, wheels, knobs, etc. that the subject manipulates; but these too are easily replaced.

Stabilization and task definition

Despite these many advantages psychologists have not rushed to study the new games or use them in prediction and performance testing. The first studies of programmable video games from a psychological standpoint were begun in the late summer of 1978 at the Navy Aerospace Medical Research Laboratory (NAMRL) in New Orleans. The purpose of the NAMRL studies was to find out whether or not the video games were suitable for inclusion in a performance test battery for environmental research.

A prime requirement of any performance test is that it stabilize. In a good performance test there comes a point in practice after which individual performance does not change in the absence of external changes. In group terms the mean follows a flat course, the variance among subjects remains the same from one trial to the next, and all correlations among stabilized trials are equal except for sampling variations. If a test satisfies these requirements, it may be used to study the impact of environmental variations on performance. If it does not, it is at best difficult to determine whether an observed change in performance is a practice effect or the result of environmental changes. An additional

requirement is that task definition (the average correlation among stabilized trials) be high, preferably greater than .90.

In the New Orleans laboratory a large number of conventional tests and, after September, 1978, video games have been studied over extended periods of practice, 15 consecutive working days, with a view to finding out how quickly, if at all, they stabilize and how well defined they are. So far nine video games have been studied in small samples (roughly 13 subjects) and one game, Air Combat Maneuvering (ACM), has been studied in roughly twice that number of subjects. All ten video games are made by Atari.

ACM is a remarkable task. The mean follows a classical learning curve, rising rapidly in the early trials and then gradually flattening out. The variance among subjects stabilizes after day 8 and the intertrial correlation after day 6. Task definition is very high, .93. In the first six days of practice, that is, prior to stabilization, the intertrial correlations show an exceptionally regular superdiagonal form. Altogether ACM not only meets the requirements laid down for it as a performance test but does so more fully than any conventional test, with one exception, studied at NAMRL. The exception is Arithmetic, a conventional test that seems to be stable from the outset; the reason, in all probability, is that arithmetical skills have been so thoroughly practiced in school and everyday life that the subjects come to the laboratory at or near asymptotic levels.

Data concerning other video games studied at NAMRL are more preliminary. It does seem, however, that some other games are as promising for performance testing as ACM. Breakout, for example, seems also to stabilize after six days, though with poor task definition, .77. It also seems that video games do not all depend on the same underlying skills and abilities since the correlations between tasks are in some cases quite low.

Convergence-divergence relations

The present report focuses on convergence-divergence relations among video games. When a task is practiced, its correlation with an external measure may increase, decrease, or remain the same, to take linear possibilities only into account. If the correlation increases, the task is said to converge on the external measure; if it decreases, the task diverges from the external measure.

Table 1 presents the cross-correlations between ACM and Breakout in 13 Navy enlisted volunteers. Each subject played 10 games of ACM a day for 15 consecutive working days, followed by 10 games of Breakout a day for another 15 consecutive working days. His score each day was the average of the 10 games played.

Now consider the row averages. These figures represent the correlation between each of the 15 days on ACM with the 15 days on Breakout considered as a whole. Testing for linear trend in a two-way analysis of variance, using the interaction between rows and columns as the error term, shows a small but significant tendency ($p < .01$) for ACM to converge on Breakout. The regression line rises by .07 from day 1 to day 15. Breakout, on the other hand, converges strongly on ACM. The regression line for the column averages rises by .33 from day 1 to day 15.

Two points are worth underscoring. First, convergence-divergence relations are not symmetrical. Because task A converges on task B it does not follow that task B converges on task A; task B may, in fact, diverge from task A. Second, Breakcut followed ACM in time. Therefore, the correlations between Breakout and ACM increased with increasing temporal separation. Day 1 on Breakout followed ACM directly while day 15 came almost three weeks later. Nevertheless, the correlations with ACM increased systematically over this interval. This result is without precedent in the literature of differential psychology; in all other studies the correlation between the same or similar measures either decreases with increasing temporal separation or remains the same.

ACM and Breakout were the first two in a series of five video tasks; the other tasks were, in order, Surround, Race Car, and Slalom. The same 13 subjects practiced all five tasks. Breakout converges strongly not only on ACM but on the other three tasks as well; linear change over the 15 days is roughly the same in all four cases, on the order of .30. ACM, however, shows no change with Surround, a slight but significant divergence from Race Car and a stronger divergence from Slalom. The linear decrease from day 1 to day 15 is .06 for Race Car and .13 for Slalom. The last two cases are the obverse of the relations between ACM and Breakout. ACM precedes Race Car and Slalom. Therefore, since it diverges from these two tasks, the correlation between ACM and Race Car or Slalom decreases as ACM gets closer and closer temporally and sequentially to the two following tasks. These results are also without precedent in the differential literature.

Application to pilot selection

A test converges on or diverges from a training criterion according as the correlation between test and criterion increases or decreases with practice on the test. If the test diverges, there is plainly no point in extending practice on the test since the effect is to lower predictive validity; if it converges, however, there may be no predictive validity at all without extended practice.

Pilot training takes place in a series of stages, each one (except the first) building on at least some of the preceding stages. It is possible, therefore, to speak not only of a test converging on or diverging from the criterion but also of the criterion converging on or diverging from a test. If the correlation between flight grades, for exmaple, and a test increases with level of training, the criterion converges on the test. If the correlation decreases as students progress to more and more advanced stages, the criterion diverges from the test. In the first case, where training criteria converge on a test, we have reason to believe that the test will predict operational performance at least as well as it does performance in training. If the training criterion diverges from a test, however, the test may easily be valid in training but much less so or not at all in operations.

TABLE 1

Cross-correlations between Air Combat Maneuvering (ACM) and Breakout in 13 Navy volunteers

	Breakout															
ACM	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	$\bar{R}$
1	29*	34	33	58	38	47	68	73	62	62	79	65	72	62	83	58
2	12	42	54	56	43	57	82	77	64	83	71	56	85	78	83	63
3	00	26	43	47	36	49	76	65	51	73	70	47	76	64	79	53
4	26	36	59	53	42	52	65	68	60	61	64	50	75	68	70	57
5	28	43	58	64	59	66	78	66	61	73	71	65	73	78	82	64
6	34	45	50	62	51	61	76	79	59	65	68	65	70	67	81	62
7	28	37	54	54	47	57	70	64	53	66	63	57	69	72	73	58
8	29	40	50	61	48	56	78	74	61	67	71	64	78	75	81	62
9	43	53	59	63	56	72	68	81	66	67	73	69	66	61	79	65
10	28	40	55	54	52	63	73	66	51	62	63	60	68	64	73	58
11	25	34	51	48	41	53	70	65	47	61	63	54	69	61	72	54
12	34	46	53	56	54	67	73	73	60	66	67	61	66	67	73	61
13	40	53	72	63	62	73	74	69	62	68	68	65	64	72	73	65
14	47	52	61	73	61	65	75	74	64	64	71	72	70	75	82	67
15	40	51	63	66	59	70	78	74	65	71	77	72	72	72	82	67
$\bar{R}$	30	42	54	59	50	61	74	71	59	67	69	61	72	69	78	61

*Decimal points omitted.

The Effect of Cognitive Style on Performance of A Critical Tracking Task

T. R. Mabry, D. A. Harris and G. A. Berry
United States Air Force Academy

Abstract

The effects of cognitive style, as determined by a set of perceptual-cognitive tests, on the performance of a visual tracking task were investigated. Subjects attempted to control a first-order unstable system with a continuously decreasing time constant. Subjects in the field independent profile performed the tracking task better than subjects with a field dependent profile. However, the difference was not statistically significant ($P = .078$). Results suggest that the profile approach to classification offers advantages over previously used multiple regression equations.

The concept of cognitive styles has experienced recent interest in various areas of psychology. In general, cognitive styles are viewed as characteristic, self-consistent modes of functioning which an individual shows in perceptual and intellectual activities (Witkin, Oltman, Raskin & Karp, 1971). The emphasis in research on cognitive style has been on shared "styles" and "style profiles" and their appropriateness for various situations and tasks.

Perhaps the earliest work in the area of such styles was really more involved with the aspect of "perceptual style." A review of perceptual literature by Long (1974) cites the following examples of perceptual styles: "confident" versus "cautious" perceivers, "levelers" versus "sharpeners," "repressors" versus "sensitizers," "high" versus "low" perceivers, "augmenters" versus "reducers," "narrow" versus "broad" band-pass individuals, and the most currently popular of perceptual style dimensions, "field dependency-independency." All of these both relate to the perceptual tendencies of a person and pertain to not just a single characteristic or trait, but to the person's general orientation in all situations. In addition, each attempts to divide the population into groups not only polarized on the perceptual dimension, but in many other consistent and predictable ways as well.

A May 1979 research report by the Air Force Human Resources Laboratory evaluated most of the above styles as well as several others. They concluded that the style dimension of field dependence-independence seems best suited for research into the intervention between styles and training. This dimension was viewed as well researched, consistent, based on well defined constructs, and accurately measurable in comparison with the other style dimensions. Both ends of the field dependence-independence continuum were viewed as holding abilities with considerable implications for training and instruction.

Long (1974) pointed out the relation between cognitive style and psychomotor performance, noting in particular that field independency is extremely frequent among pilots. He also chronicled the relation between gender and cognitive style; i.e., females tend to be more field dependent and males more field independent. It is important to note, however, that field dependence-independence is not a dichotomy, but rather a continuum.

PRECEDING PAGE BLANK-NOT FILMED

Therefore, persons may differ to the extent their psychomotor performance is affected by their perceptual style. Measures of other cognitive dimensions which are related to psychomotor abilities should contribute to prediction of psychomotor task performance when used in combination with a measure of field dependence-independence.

Since style difference (field dependence-independence) has been reported to be related to various behavioral phenomena such as information processing and problem solving ability (Long, 1974), measures of such abilities might prove meaningful additions to improve prediction of psychomotor performance. Therefore, profiles of cognitive/perceptual measures including not only field dependence-independence, but also measures of abilities such as perceptual speed, visual memory, spatial orientation, etc., and gender might provide meaningful groups to compare on psychomotor performance of a highly discriminatory task.

Previous attempts to validate findings based on regression approaches to predicting psychomotor task performance have failed to find consistent prediction from sample to sample and between sexes (Becker, Williges, Williges & Koonce, 1979). Another approach is to form groups based on the multivariate similarity of individual profiles to a prototype vector of scores on the previously discussed variables (Harris, 1976). The current study is an attempt to relate cognitive style and gender to the performance of a visual tracking task.

Method

Subjects. Some of the subjects for this investigation were 16 paid individuals (8 male and 8 female) from the subject pool at NASA Ames Research Center. The remaining 41 subjects (33 male and 8 female) were volunteers from psychology classes at the United States Air Force Academy.

Test Apparatus. A Critical Task Tester (CTT) MKII was used to administer the visual tracking test (Jex, McDonnell, & Phatak, 1966). The CTT includes a CRT display of a luminous horizontal line which is displaced vertically proportional to the tracking error and indicates the value of the unstable root λ_c at the end of each test run. The display unit is connected to a master control unit which provides basic electronic circuitry, power supply and the experimenter's control unit. Displays on the experimenter's unit provide status information about the state of the system and the score for the most recently completed trial.

The subject controls the location of the horizontal line with vertical movements of a fingertip control. The stick is very nearly a force stick and the deflection required is quite small. The control stick is connected to other units of the CTT through the CRT Display.

To measure the effective time delay, the test apparatus is operated in the autopaced mode. The value of the system time constant at the end of a trial is the index of the subject's effective time delay for that trial. Thus the value of the time constant at the point where control is lost is a measure of the subject's tracking ability with a given display and control device. Jex and Allen (1970) have demonstrated that the inverse of this critical root (λ_c) is a sensitive measure of performance. At the start of a trial the system automatically set the time constant at zero, but linearly increased rapidly over time as long as good control was maintained. When the error was 10% of the maximum allowable value, the rate was reduced by a factor of four. Loss of control was defined as the

display system error going off scale and the resulting dependent measure was the level of the critical root attained at the trial's conclusion. Each subject performed the task in a darkened, sound attenuated chamber. The subjects were seated in front of the CRT display so that they could comfortably operate the control stick.

Procedure. During the first experimental session, the subjects were given a battery of five cognitive tests. In addition to the cognitive tests, the subjects were given 30 trials on the CCT in two 15 trial blocks with a two minute rest between blocks. In the second experimental session the subjects were given 45 trials on the CCT. The trials were again administered in three 15 trial blocks with a two minute rest period between blocks.

Data. The experimenter recorded the critical root for each of the seventy-five CCT trials. Medians for each subject for each 15 trial block were then calculated. The measure of tracking performance used in this investigation was the median critical root for the last block of trials (i.e., trials 61 through 75).

The cognitive style data were taken from a battery consisting of five tests. The test battery included: (a) identical pictures test (perceptual speed), (b) map memory test (visual memory), (c) cube comparison test (spatial orientation), (d) maze tracing test (spatial scanning), and (e) Embedded Figures Test (field dependence-independence). The Embedded Figures Test is from the Educational Testing Service (Witkin, Oltman, Raskin & Karp, 1971) and the first four are from the Ekstrom, French, Harman and Derman (1976) battery. These same tests were previously administered to 48 males and 48 females for the purpose of establishing score profiles.

Data Analysis and Results

Cognitive Style Profiles. The method used to develop the cognitive style profiles was linear typal analysis (Overall & Klett, 1972). The aims of the analysis were to discover the nature and number of underlying cognitive styles and to define a prototype vector to represent each cognitive style. Each subject was then assigned to the cognitive style profile most similar to his or her own. Table 1 contains the results of the linear typal analysis for the 96 subjects in the profile development group. Of the 57 subjects in this study, 26 were placed in profile I (called field independent) and 31 were placed in profile II (field dependent).

TABLE 1

Means for Cognitive Style Profiles (n = 96)

	Profile I	Profile II
Identical Pictures	38.45	34.93
Map Memory	11.07	10.82
Cube Comparison	14.37	7.62
Maze Tracing	16.50	14.22
Embedded Figures Test	21.89	40.66**

**mean sec/item

Analysis of Variance. A three-way (sample x gender x profile) analysis of variance was performed on the last block of 15-trial median values of λ_c. There was a significant sample by gender interaction, $F(1,44) = 7.506$, $p < .01$ and a significant sample by profile interaction, $F(1,44) = 5.276$, $p < .05$. The triple interaction (sample x gender x profile) was also significant, $F(1,44) = 8.125$, $p < .01$. Tracking performance was not significantly affected by any of the three main effects. However, the effect of cognitive style on tracking performance approached significance, $F(1,44) = 3.248$, $p = .078$.

Conclusions

Based on the results of this study, the cognitive style classification technique yielded consistent classifications from sample to sample. However, there were a few cases in each sample that were not clearly classified into one profile or the other. These outliers could have contributed to the non significance of the profile effect. Since this was an exploratory study, the outliers will be removed and the data will be reanalyzed. The interactions between gender and sample, profile and sample, and the triple interaction all could be due to the atypical, homogeneous sample at the United States Air Force Academy. Continuing research is in progress in an attempt to better predict performance of complex psychomotor tasks. These exploratory results, however, appear to indicate the possible value of style profiles as predictors of task performance.

References

Becker, R. J., Williges, B. H., Williges, R. C. and Koonce, J. M. Prediction of performance in motor skills training. Proceedings of the 23rd Annual Meeting of the Human Factors Society, 1979, 300.

Ekstrom, R. B., French, J. W., Harmon, H. H., and Dermen, D. Manual for kit of factor-referenced cognitive tests. Princeton, N. J.: Educational Testing Service, 1976.

Harris, D. A. Classification typology for predicting performance in Air Force technical training (AFHRL Tech. Rep. 76-91). Lowry Air Force Base, CO: Air Force Human Resources Laboratory, Dec., 1976.

Long, G. M. Reported correlates of perceptual style: A review of the field dependence-independency dimension. JSAS Catalog of Selected Documents in Psychology, 1974, 4, MS No. 1083.

Jex, H. R., McDonnell, J. D., and Phatak, A. V. A "critical" tracking task for manual control research. IEEE Transactions on Human Factors in Electronics, 1966, 7. 138-145.

Jex, H. R. and Allen, R. W. Research on a new human dynamic response test battery. Part I: test development and validation; Part II: Psychophysical correlates. Sixth Annual NASA-University Conference on Manual Control, Wright-Patterson AFB, OH, 7-9 April 1970, 743-777.

Ragan, T. J., Back, K. T., Stansell, V., Ausburn, L. J., Ausburn, F. B., Butler, P. A. and Huckabay, K. Cognitive styles: A review of the literature (AFHRL Tech. Rep 78-90-1). Brooks Air Force Base, TX:

Air Force Human Resources Laboratory, May 1979.

Witkin, H. A., Oltman, P. K., Raskin, E. and Karp, S. A. A manual for the Embedded Figures Tests. Palo Alto: Consulting Psychologists Press, 1971.

Validation of a Proposed Pilot Trainee Selection System

Jefferson M. Koonce
United States Air Force Academy, Colorado

Since the Spring of 1976 there has been an emphasis on the need to implement a ground-based screening (GBS) program for the selection of pilot training candidates, to include all sources of entry to pilot training: AFROTC, OTS, and the Academy. The program that was urged to be implemented consisted of the Air Force Officer Qualifying Test and an aircrew psychomotor test (APT) device which presented two tasks: a two-hand pursuit rotor task (complex coordination) and a stick-and-rudder task (similar to the Mashburn task of 1934). In 1976 claims of significant validity coefficients had been made regarding GBS, but there was no evidence of the predictive validity of the program; that is, a relationship between the predicted success/failure of candidates and their actual performance in pilot training. Furthermore, because of the highly selected nature of our cadet population we questioned the validity of the GBS program if it were used to screen Academy cadets for pilot training.

After numerous meetings and correspondence about our concern that the program had not been validated, we decided to make our own validation check of the proposed ground-based screening program in the Fall of 1977.

Design of Validation Study at USAFA

Class	1978	1979	1983
Sample Size	129	289	798
AFOQT	Dec 77	Dec 77	Jul 79
APT	Spring 78	Spring 78	Summer 79
Graduate	May 78	May 79	May 83
Complete Pilot Trng	Jan 80	Jan 81	Jan 85

Copies of the AFOQT form N and several of the psychomotor test devices were obtained from the Air Force Human Resources Laboratory (AFHRL), Brooks AFB, Texas. We tested random samples of the Class of 78 and 79. Of the class of 83, all took the AFOQT and almost all of the pilot qualified cadets took the psychomotor test. The cadets in the Class of 78 had completed the T-41 (light aircraft) training program before taking the GBS tests, and the other two groups had not taken T-41 before the GBS tests.

The AFOQT was scored by AFHRL and a copy of the data given to the Academy. The data on the Class of 83 has not been received at this time.

AFOQT	1978	1979
Mean	119.070	111.685
s.d.	47.400	50.336
Passed	116	159
Failed	13	22

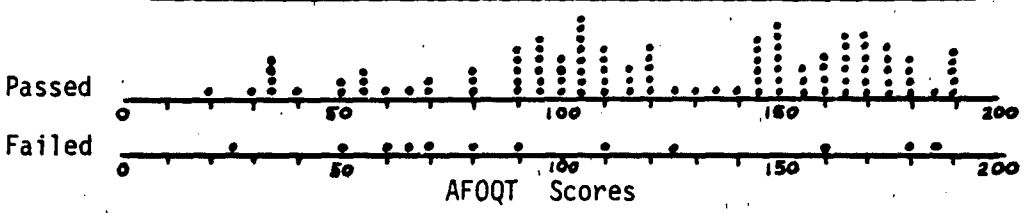
The AFOQT criteria for acceptance to pilot training is at least the 25th centile on the pilot composite, the 10th centile on the navigator composite and at least a total of 50 when combining the pilot and navigator centile scores. The subject data used to calculate the statistics in the above table were the sums of the pilot and navigator centile scores.

AFHRL has recommended that only the second psychomotor test be used in the GBS program because the first test does not contribute significantly to the validity of GBS. They also provided the mean and standard deviation for each of the three axes scored on the second test. Each subject's scores were converted to z-scores using the data provided by AFHRL, and the three z-scores were summed. The AFHRL suggested criterion for selection for pilot training was a total of the z-scores less than +3.00. Thus, if a person's error was one standard deviation above the mean on all three axes he would not be permitted to go to pilot training.

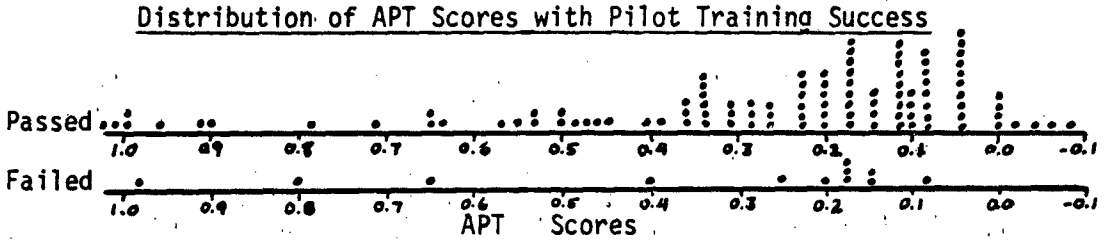
APT	1978	1979
Mean	-0.395	-0.407
s.d.	2.571	2.164
Passed	116	239
Failed	13	22

At the present time, only the data on the Class of 78 has matured, and it is presented below.

Distribution of AFOQT Scores with Pilot Training Success



Distribution of APT Scores with Pilot Training Success



Because the AFOQT and APT do not seem to be able to adequately identify those cadets that would have been unsuccessful in pilot training, we are looking for other means of predicting cadet success in flight training. Currently, we are investigating various cognitive as well as personality measures as predictors, and some of those results will be presented in a later session. Another potential predictor being researched at the Academy is the rate at which one can learn complex flight maneuvers, such as Chandelles and Lazy Eights, using a desk-top flight simulator. This research thrust is particularly interesting because, if it proves adequate for selection, it would also provide some valuable training of some flight skills that would transfer to the pilot training of the individual.

The Pilots' Night Training Device--An Answer to the Army's Night Terrain Flight Training Enigma?

Colin D. Ciley
Canyon Research Group, Incorporated
Westlake Village, California 91361

Martin F. Allnutt
British Exchange Scientist
U.S. Army Research Institute Field Unit
Fort Rucker, Alabama 36362

Abstract

Various constraints limit the Army's ability to train its pilots in essential night flight at very low altitude. This paper reports a method of circumventing some of them. Sixteen Army aviators were tested (T_1) at night on their performance of essential terrain flight maneuvers. After being divided into two matched groups, one group trained in the daylight while wearing a specially designed light attenuating device; the other group received the same training, but at night. Retesting at night (T_2) showed that the performance of both groups improved significantly ($p < 0.01$) between T_1 and T_2; furthermore, the group trained by day performed better than the group trained at night. Potential training implications and further applications of the technique are discussed.

INTRODUCTION

The Problem. The Army's concept of the employment of helicopters in combat requires night terrain flight as a principal means of increasing effectiveness. However, unit training in night terrain flying techniques is and has been difficult, sometimes virtually impossible, to conduct. Safety considerations, shortage of unit instructor pilots capable of training students to perform night terrain flight missions, disruption to circadian rhythms, and perhaps most importantly, local, civil restrictions on low level night flight, all act to constrain or deny effective unit training in these essential skills. Simulation of night visual conditions through attenuation of daytime ambient light has been under investigation for over thirty years; it may yet provide an acceptable solution to this enigma.

Early Work. Research in attenuation of light to simulate the night visual condition has been conducted, albeit sporadically, at least since the late 1940s. Scientists of the Army Research Institute (ARI) have been active in this area since 1973. They have investigated technical specifications for devices based on the technology as well as possible applications of such devices to Army training. An excellent overview of this work has been prepared (Bleda and Farrell, 1979).

Preliminary Work at Fort Rucker. In the spring of 1978, a set of Neutral Density (ND) 5.1 filters prepared earlier for testing in the Army M17 Protective Mask were modified to fit the standard military Sand, Wind and Dust (SWD) goggles, such that they could be used in experimental use in terrain flying training applications. They were manufactured by Omnitech,

PRECEDING PAGE BLANK-NOT FILMED

Inc., of Dudley, MA, who also provided five auxiliary plastic filters having "fractional" optical densities which ranged from ND 0.56 to ND 0.91. The principal objectives of the preliminary investigation of this design were to verify the optical density recommendations found in the literature and to determine the acceptability of the concept by aviators and instructor pilots. Flights using experienced aviators and a range of filter densities were conducted under weather conditions ranging from clear, bright sunlight to overcast, with low cloud and heavy rain. The aviators were unanimous in their opinion that the visual experience using the goggles was essentially the same as the experience of flying at night and were supportive of the concept.

As these experiments continued, it became increasingly apparent that an optical density of ND 5.1 was not "dark" enough, even on overcast days. On "bright" days, the scene was that of an unrealistically bright moonlight night. Accordingly, a further study of the light parameters involved (e.g., illuminance, luminance, reflectance) was conducted to select an optimal density or range of densities which might give more realistic conditions and operationally usable results.

As a result of these investigations, specifications for the prototype Pilots' Night Training Device (PNTD), a specially modified standard pair of Sand, Wind and Dust Goggles, with a choice of two filters, ND 6.0 and ND 7.0, were developed by the researchers. The ND 6.0 filters were expected to provide the visual experience of various nighttime conditions, from a bright moonlit night to a dark starlit one, under commonly experienced daylight illuminance conditions. The ND 7.0 filter was selected to provide the visual experience of a one-half moon or less night "scene" when the filter is used under extremely bright daylight conditions ($1 - 1.13 \times 10^4$ foot candles illuminance).

These lenses, also prepared by Omnitech, Inc., were made from dyed and molded polycarbonate. The goggle frame was modified at Fort Rucker to provide an airflow-induced vacuum inside the device, thereby providing a means for evacuating the warm, moist air and aiding in reducing condensation. The prototype device was ready for testing early in 1979, and an experiment to determine its efficacy and any problems pertaining to its use began at the US Army Aviation School, Fort Rucker, shortly afterwards. An account of this experiment forms the remainder of this report.

METHOD

Subjects. Sixteen (16) recently graduated UH-1 Instructor Pilots (IPs) with flight time experience averaging 1400 hours (range 600-2800) served as subjects. They were all able to dark adapt satisfactorily and none of them had received previous Army school training nor intensive field experience in night terrain flight.

Test Personnel. The testing and training aspects of the experiment were kept entirely separate. Two very experienced Evaluators (night flight Standardization Instructor Pilots) conducted the pretests and posttests, and two experienced night flight IPs conducted the training. During pretest and posttest flights, scientist/pilots assisted each Evaluator by acting as Recorders. Prior to the start of the experiment, the Evaluators, Instructors and Recorders flew together several times to refine and standardize their judgments and procedures.

Pilots' Night Training Device (PNTD). The device was as described earlier.

Testing and Training Areas. Testing and training flights were conducted in separate, though similar, training areas near Fort Rucker. Hovering maneuvers took place in a large pasture, while terrain flights were conducted over rolling, heavily-forrested countryside.

Experimental Design. All subjects were pretested (T_1) at night on basic standard hovering and terrain flight maneuvers and divided, according to their scores, into two matched groups. One group (A) then received five ninety-minute nighttime training sessions (TR 1-5), while the other group (B) received five identical training sessions by day while wearing the PNTD. In order to balance any delays due to inclement weather and maintain identical spacing of training, Group A completed their training first, on consecutive nights, and then the spacing of Group B's training was adjusted to ("yoked" to) that of Group A.

Procedure. Before each test or training session, subjects were briefed and dark adapted. Each session included 30 minutes in transit, 40 minutes on hovering maneuvers, and 20 minutes on terrain flight. The maneuvers were sequenced in order of increasing difficulty, as they are usually taught, and were always carried out in the same order. Training sessions consisted of demonstrations, practice and instructor feedback.

When they had completed their training periods, both groups returned to the pretest area and again were tested on their flying performance at night (T_2). Pretest and posttest measures, as well as the test area used, were identical for both groups.

Table 1

Design of the PNTD Experiment

Pretest	Groups	Training Periods					Posttest
T1	Group A Night Trained N = 8	1	2	3	4	5	T2
	Group B PNTD Trained N = 7*	1	2	3	4	5	

(* One subject withdrew for medical reasons)

The date for T_2 was chosen so as to match the moon phase and altitude conditions which occurred during T_1 and to ensure that all test and training sessions were conducted under similar illumination conditions. All tests were conducted under half-moon conditions with the moon at least thirty degrees above the horizon. Illuminance was measured and recorded in each occasion.

Performance Assessment. Subjects were tested and trained on nine basic hovering maneuvers and on terrain flight over a 20 km course. During T_1 and T_2, performance on each maneuver was assessed by the Evaluator who estimated the variation from criterion on various parameters. Such variations were coded on a gradeslip which was specially developed for this

experiment and subsequently converted into a composite score for each maneuver. The Evaluator reported his measurements to the Recorder in code in order to avoid giving distracting feedback to the subject. The Recorder also noted maximum radar altimeter readings (skid height) during each hovering maneuver and at standardized checkpoints along the terrain flight course. Thus, although the procedure still contained a large subjective element, every effort was made to increase objectivity.

Additional measures taken included pre- and post-experiment questionnaires on the subjects' experience and attitudes. Illuminance and weather data for each testing and training session were recorded.

RESULTS

Performance Scores. As can be seen from Table 2, the performance of both groups improved markedly from T_1 to T_2; this improvement was statistically significant ($p < 0.01$) for both groups. Trends in the data indicated that Group B improved more from T_1 to T_2 than Group A, and that Group B's performance on T_2 was better than that of Group A on T_2; however, neither of these results reached statistical significance ($p > 0.05$). The scoring system established that the better the performance, the lower the score.

Table 2

Ranges, Means, and Standard Deviations of Hovering Maneuver Performance Scores of Groups A and B on the Pretest (T_1) and Posttest (T_2)

	T_1			T_2		
Group	Range	X	σ	Range	X	σ
Group A (N = 8) (Control)	11 - 30	23.4	5.5	4 - 21	11.8	5.0
Group B (N = 7) (PNTD)	13 - 32	22.9	7.9	5 - 14	8.7	3.6

As indicated in Table 2, the relationship between mean score and the range of individual scores showed that the distribution of performance scores was relatively symmetrical.

Skid Heights. A summary of the hover maneuver skid height data (radar altimeter readings) is presented in Table 3. Both groups performed better, i.e., lower skid heights, on T_2 than on T_1; Group B exhibited a greater improvement between T_1 and T_2 than did Group A. However, none of these differences was statistically significant ($p > 0.05$).

Table 3

Ranges, Means and Standard Deviations, in feet, of Hovering Maneuver Skid Heights, of Groups A and B on the Pretest (T_1) and Posttest (T_2)

	T_1			T_2		
Group	Range	X	σ	Range	X	σ
Group A (N = 8) (Control)	3.8 - 8.2	6.4	1.4	3.8 - 6.8	5.3	1.0
Group B (N = 7) (PNTD)	4.5 - 12.3	6.7	2.4	2.6 - 6.5	4.5	1.2

Questionnaire Data. The subjects who trained with the PNTD stated that it enhanced their capability to practice night terrain flight, that they felt safer and more relaxed than when flying in actual night conditions, that the feedback from their instructors was more precise than in their previous experience at night, and that the "night scene" was realistic. All experienced some condensation problems and most expressed a desire to have visual access to some of the aircraft instruments.

DISCUSSION

The finding that training with the PNTD appears to be at least as effective as training at night is encouraging. Some caution is advised, however, in view of the limited number of subjects and maneuvers used in this experiment, the limited amount of training time and the inevitable subjective element in the perforamnce assessment. However, a very important finding is that the pilot techniques learned with the PNTD do transfer immediately and well to true night conditions. Of almost equal importance is the likelihood that the instructor pilot is relieved of much of the night-imposed stress of his safety pilot duties, and thereby can concentrate on providing more complete and accurate feedback to his student. Lastly, the subjects' response to the PNTD was very positive; an important asset for any potential training device.

CONCLUSIONS

The use of the PNTD in daytime is a viable alternative to conventional night-time training in terrain flight maneuvers. The advantages it offers in safety, convenience, and relief from environmental objections to actual night flight is a strong argument for continuing work in this area.

FURTHER DEVELOPMENTS

Currently, the use of the PNTD as a means of training nighttime aircraft emergency procedures and night terrain flight navigation is being investigated. This is another step towards the ultimate objective of the research, the development of a complete night terrain flight training program unconstrained by present night flying training restrictions.

REFERENCE

Bleda, P. R. and Farrell, J. P. Development of Light-Attenuating Devices (LADS) to Simulate Night Visibility During Daylight. U.S. Army Research Institute Technical Paper 375. July 1979.

"The views, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other official documentation."

GLIDESLOPE DESCENT RATE CUING

TO AID CARRIER LANDINGS

By

Gavan Lintern, Canyon Research Group, Inc., Orlando, FL
LT Charles E. Kaul, USN, Naval Air Engineering Center, Lakehurst, NJ
Stanley C. Collyer, Naval Training Equipment Center, Orlando, FL

ABSTRACT

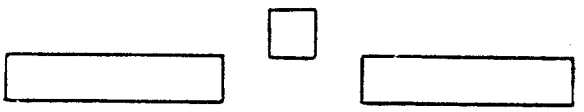
The Fresnel Lens Optical Landing System (FLOLS) used for glideslope guidance during aircraft carrier approaches, provides zero-order or displacement error information for the pilot to judge whether he is above or below the glideslope. The higher-order components in the aircraft control system create substantial lags between inappropriate control behavior and subsequent error indications from the FLOLS. Landing performance of experienced Navy pilots was tested in the Navy's Visual Technology Research Simulator (VTRS) with an augmented FLOLS display, in which variable-length vertical bars provided first-order information for glideslope guidance. Two algorithms to drive the augmented display were compared with a conventional FLOLS. One, termed RATE, provided information on the rate of displacement of the aircraft from the glideslope. The other, termed COMMAND, additionally incorporated a desired return rate (dependent on displacement from the glideslope) into the computation.

Both first-order displays reliably improved glideslope tracking performance throughout the approach. Lineup performance was not adversely affected. Any differences between the two first-order configurations favored the COMMAND display. The pilot subjects were unanimous in strongly endorsing the modified systems and preferred the COMMAND to the RATE display. If comparable improvements in glideslope performance are found in carrier-based operations, boarding rates and glideslope-related accident rates can be expected to improve substantially.

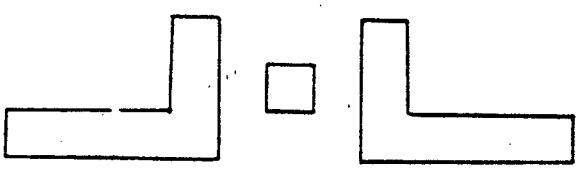
INTRODUCTION

The FLOLS provides primary glideslope information for a pilot approaching to land on an aircraft carrier. It consists of light sources behind five vertically stacked Fresnel lenses situated between two horizontal light arrays known as the datum bars. The light is visible to the pilot through the center lens when he is within 9.5 minutes of arc of the glideslope, and is seen as level with the datum bars. As the aircraft moves more than 9.5 minutes of arc above or below the glideslope, the light is seen through higher or lower Fresnel lenses to give the appearance of a single light moving vertically above or below the line of the datum bars. The apparently moving light is known as the meatball.

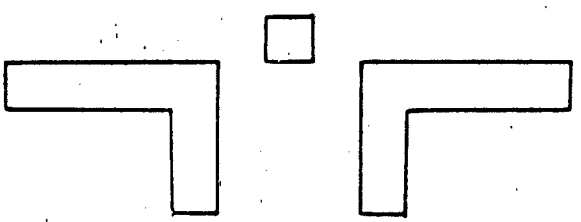
PRECEDING PAGE BLANK-NOT FILMED



(a) A static CONVENTIONAL display does not permit a trend interpretation. This state for the RATE display, indicates that the one-ball high condition will be maintained, while for the COMMAND display that the pilot is returning to the reference glideslope at an appropriate rate.



(b) This state for the RATE display, indicates one-ball high and going higher in relation to the reference glideslope. For the COMMAND display, this state indicates that the aircraft is high, and is not returning to the glideslope quickly enough.



(c) This state indicates for a RATE display that the pilot is returning to the glideslope, while for the COMMAND display that he is returning to it too quickly and will probably fly through it.

Figure 1. Three representations of possible RATE or COMMAND display states. Figure 1(a) can also represent a CONVENTIONAL display state.

The displacement information from the FLOLS is helpful for glideslope control but is less than optimum. Because the information from the meatball is of zero-order, there are substantial lags between incorrect control inputs and the subsequent error information from the FLOLS. For example, a rate (first-order) error must exist for some short period of time before it produces a perceptible displacement (zero-order) error. The pilot is more directly in control of rate than of displacement, so that he could correct the rate error if he were aware of it before a substantial displacement error was incurred.

For this experiment vertical bars of variable length could extend up or down from the inside ends of the datum bars (Figure 1). For the RATE display algorithm, the length of the bars was proportional to the rate of displacement of the aircraft above or below the glideslope, with bars extending down to indicate a downward angular displacement of the aircraft relative to the glideslope. For the COMMAND display algorithm, the bar length was proportional to the difference between the displacement rate

described above and a commanded rate that was a function of aircraft displacement from the glideslope. For this display, the vertical bars extended up when descent rate was insufficient and down when it was excessive. (Refer to Kaul, Collyer, and Lintern, in press, for a detailed description of these algorithms.)

METHOD

Eight experienced carrier-qualified Navy pilots landed a simulated T-2C aircraft on a representation of the Forrestal aircraft carrier with the CONVENTIONAL, RATE, or COMMAND FLOLS displays. The Naval Training Equipment Center's VTRS at Orlando, Florida, was used for the experiment.

APPARATUS. The VTRS has a fully instrumented T-2C Navy jet-trainer cockpit, a six-degree-of-freedom synergistic motion platform, a 32-element G-seat, a wide angle visual system that can project both computer-generated and model-board images, and an Experimenter/Operator Control Station (Collyer and Chambers, 1978). The motion system, G-seat and model-board were not used in this experiment.

Visual System. The background subtended 50^{0} above to 30^{0} below the pilots' eye level, and 80^{0} to either side of the cockpit. The carrier image was generated by computer and projected onto the background through a 1025-line monochrome video system.

Simulation Configuration. The simulator was initialized with the aircraft at 9000 feet from the carrier, on the glideslope and centerline, and in the approach attitude and configuration (hook and wheels down, speed brake out, 15 units angle of attack, and power at 83%). The T-2C is normally landed with full flaps, but flaps were set at half extension for this experiment to more closely simulate approach speeds of typical fleet aircraft. Fuel was set at 1320 lbs to give 10,000 lbs gross weight. A landing trial was flown from the initial condition to touchdown on the carrier or, in the case of a bolter (touchdown beyond the wires) or waveoff, to 1000 feet past the carrier. The carrier was set on a heading of 360^{0} at 20 knots. Environmental wind was set at 317^{0} with a velocity of 6.34 knots. This combination of carrier speed and environmental wind produced a relative wind component of 20 knots directly down the landing deck.

Difficulty Variables. Time of Day and Turbulence factors were included to examine the effects of the FLOLS displays under a variety of flying conditions. Time of Day was manipulated with day and night computer-generated data bases (having identical FLOLS displays). The day scene included a horizon, a solid-surface carrier image, and a 2000 ft. wake. The night scene had a featureless background, no wake, and a carrier image consisting of point lights depicting the landing deck, centerline and drop lights. Turbulence, present on half the trials, consisted of a random buffeting in the vertical dimension.

PROCEDURE. Pilots flew a practice block of 20 approaches and three experimental blocks of 34 approaches over three days. Only one FLOLS type

was used in any 34-trial block, and each pilot flew with each of the three FLOLS types. The order of testing the three display conditions was counterbalanced across subjects. The first 10 trials of each 34-trial block were used to practice approaches with the relevant FLOLS type. The independent variables of Turbulence and Time of Day were fully crossed with FLOLS type. The design counterbalanced for trends across and within blocks, and used buffer trials to counter possible interference between experimental conditions.

RESULTS

Table 1 summarizes analyses of variance on RMS (Root Mean Square) deviations from the glideslope, performed separately on four segments of the approach. Significance levels of the main effects and interactions are presented along with η^2 (Eta Squared) values, which represent the proportion of the total variance accounted for by each factor. FLOLS type had a statistically reliable effect on glideslope performance for all segments of the approach. Interactions of FLOLS type with Turbulence and Time of Day were not statistically reliable except in two cases where the interactions accounted for less than 1% of the total variance.

Table 1. Mean RMS Glideslope Deviations: Significance Levels and Eta Squared Values (η^2) for FLOLS Type (FL), Time of Day (TI) and Turbulence (Tu) Factors and Interactions, (*=p<.05; **=p<.01)

Distance from Carrier (ft)	6000-4500		4500-3000		3000-1500		1500-0	
	Sig.	η^2	Sig.	η^2	Sig.	η^2	Sig.	η^2
FL	**	.106	**	.111	**	.124	*	.096
TI	*	.026	--	.003	--	.001	--	.001
Tu	**	.017	**	.020	**	.034	**	.085
FL x TI	--	.006	--	.005	*	.001	--	.003
FL x Tu	--	.003	--	.001	--	.010	*	.005
TI x Tu	--	.007	--	.004	--	.001	--	.006
FL x TI x Tu	--	.005	--	.001	--	.006	--	.001

Table 2. Mean RMS Glideslope Deviations (in feet, and Significance Levels of Planned Comparisons

Distance from Carrier (ft)	6000-4500	4500-3000	3000-1500	1500-0
FLOLS Type				
CONVENTIONAL (CONV)	14.74	13.83	10.83	6.09
RATE (RATE)	12.0	10.81	8.08	4.61
COMMAND (COMM)	7.84	6.98	5.27	3.67
Planned Comparisons				
COMM + RATE vs. 2 CONV	<.01	<.01	<.01	<.01
COMM vs. RATE	<.01	<.05	<.10	>.10

The mean RMS deviations for the three display conditions are presented in Table 2. It is seen that the effects are substantial and reliably favor the first-order displays. Additionally, performance with the COMMAND display was superior to that for the RATE display; this difference is statistically reliable for the two segments between 6000 and 3000 feet from the carrier.

Several measures of lineup performance were also examined to explore the hypothesis that glideslope improvement may have been gained at the expense of lineup. FLOLS display type was found to have no reliable effects on lineup performance on any segment of the approach.

In a post-experimental questionnaire, all eight of the pilots, and two Landing Signal Officers who attended the experiment during its first week to score approaches, indicated that they strongly preferred the first-order displays to the conventional FLOLS display. Nine of the ten respondents preferred the COMMAND to the RATE display.

DISCUSSION

A first-order display similar to those tested in the experiment could be provided on a carrier by integrating a relatively inexpensive optical tracker with existing hardware. With similar reductions in glideslope error at the ship as were found in the VTRS with the COMMAND display, night boarding rates (currently at approximately 90%) could be improved accordingly. Glideslope related accidents such as ramp strikes, hard landings, and in-flight engagements could also be reduced substantially from current rates. This system has considerable potential for improving efficiency at the ship and for saving aircraft and lives.

Aside from its apparent operational significance the COMMAND display might help students learn to use a conventional FLOLS display more effectively. Lintern (1980) has shown that simulator training of light aircraft landings can be enhanced with the use of supplementary visual guidance cues. For carrier landing training in a simulator the COMMAND display might similarly help students learn appropriate glideslope control techniques that then would help their subsequent performance with a conventional FLOLS display. This hypothesis will be examined in research to be conducted in the near future.

Collyer, S. C., and Chambers, W. S. AWAVS, a research facility for defining flight trainer visual requirements. Proceedings of the Human Factors Society 22nd Annual Meeting, Detroit, 1978.

Kaul, C. E., Collyer, S. C., and Lintern, G. Glideslope descent rate cuing to aid carrier landings. NAVTRAEQUIPCEN IH-322, (In Press).

Lintern, G. Transfer of landing skill after training with supplementary visual cues. Human Factors, 1980, 22, 81-88.

Perry, B. L. Optical guidance systems: analysis, design and development. Naval Research Laboratory, Report 6581, 1967.

Simulation Requirements for Helicopter Landing

G.L. Ricard
Naval Training Equipment Center
Orlando, Florida 32813

R.V. Parrish and R.R. Ashworth
NASA Langley Research Center
Hampton, Virginia 23665

Lt. M.D. Wells RN
Naval Air Test Center
Patuxent River, Maryland 20670

Abstract

Expensive subsystems are needed to enable flight simulators to produce cues of aircraft motion, and recent evidence is that these additions are not always useful for training. The Navy is assessing the needs of a trainer for the LAMPS MK III helicopter, and this work documents the effectiveness of motion cueing devices for helicopter hover operations about destroyer-class ships. Adding both a G-seat and a motion platform showed improved hover control performance relative to that seen with a fixed-base simulation. In addition, visual system delays characteristic of trainers additively degraded this performance while motion of the ship model did not affect it greatly. Pilot differences were apparent especially in the unique ways in which the pilots responded to the combinations of these conditions.

The real test of technology for training devices is whether its cost is worthwhile for training. Developments over the past ten years in visual displays for flight simulation, for instance, have extended the range of tasks for which flight simulators are useful for training. Such has not been the case for equipment designed to provide some of the sensations of aircraft motion (Puig, Harris & Ricard, 1978). Motion platforms have been useful for simulations of large commercial and military aircraft, but for the most part, they have not been shown to be cost effective devices for training systems involving high-performance, fixed-wing aircraft (Parrish, 1978), and such training constitutes a significant proportion of military flying training efforts. Generally speaking, it is far easier to demonstrate an effect of motion cues on pilot inputs to a simulator than it is to show their effect on system performance (Parrish, Houck & Martin, 1977), and to date, attempts to demonstrate the transfer effectiveness of motion platforms have been disappointing. The G-seat, a somewhat newer development, similarly has been shown to aid the performance of some tasks (McKissick, Ashworth, Parrish & Martin, 1980), but these effects are small and the likelihood of large transfer effects is also small.

A somewhat different potential exists for helicopter trainers in that the

PRECEDING PAGE BLANK-NOT FILMED

aircraf are more unstable than fixed-wing craft and often flying tasks for which synthetic training has been useful are more difficult. A good example is landing a simulated helicopter on a destroyer-class ship during high sea states. Most trainers have a limited field of view visual display, the aircraft are difficult to control, and the major visual referent (the ship) is moving. Here, being able to differentiate motion of the ship from motion of the aircraft can be very useful during the final phase of landing, and one might suspect that motion cueing devices would be more useful for simulations of this task than for simulations of fixed-wing landings. To explore this possibility, as part of an effort to specify a trainer for the LAMPS MK III helicopter, the Naval Training Equipment Center and the NASA Langley Research Center studies the effects motion cues had on helicopter hover performance.

Methods

Subjects: Twelve helicopter pilots from the Naval Air Test and Evaluation Squadron (VX-1) at the Naval Air Test Center served as subjects. They had an average of over 1780 flight hours, 1580 of them in the SH-2 and SH-3 helicopter, and they had made about 220 small ship landings.

Equipment: The device used was the Visual Motion Simulator at the NASA Langley Research Center. It is equipped with a six-post, synergistic motion platform, a second generation G-seat, and a color television model board visual display which provides a 36° by 48° field of view. A section of the model board system was used to mount a 380:1 scale model of a destroyer escort equipped with an aft landing pad and appropriate deck markings. The flight dynamics for a Huey Cobra were calculated 32 times per second, and all cueing systems - the visual display, the motion platform, and the G-seat - operated with a total system delay of approximately 70 milliseconds. The device is equipped with a generic two seat cockpit and for this study, its right-hand seat was configured for helicopter control with a McFadden two-axis controller as a cyclic and with a typical helicopter collective control. Ship motions were created by mounting the ship model on the four pads of an extra G-seat and these were appropriately driven to produce the pitching, rolling, and heaving motions of a sea state three (Fortenbaugh, 1978).

Design, Variables & Measures: The study was a factorial combination of fixed-base (FB), moving base (MB), and G-seat (G) conditions of motion cueing with two conditions of ship motion (on and off) and two system delays (70 and 133 msec) representative of current trainer technology. These twelve conditions were tested as a within-subjects, repeated measures experiment where a subject was tested five times on a given condition before proceeding to a new one. This testing was ordered such that each condition appeared once in each position. Performance was measured as system errors (lateral, longitudinal, and altitude errors as well as a vector combination of all three), plant states (aircraft pitch and roll angles), and pilot control inputs (pitch and roll inputs to the cyclic, collective inputs, and manipulations of the rudders). The datum for each measure was

its root-mean-square averaged over each one-minute trial.

Procedure & Task: Each pilot received one to two hours of familiarization flight. Initially this involved hovering over a Maltese cross on the model board's airfield, and then practice was switched to the destroyer model. To reduce the probability of damage to the TV probe, the experimental task was to hover in line with a diagonal line-up marking on the flight deck about 60 feet from the ship at an altitude of twenty feet above the deck. Each trial started with the helicopter centered in the correct position, and a computer graphics head-up image superimposed on the pilot's visual display indicated his position relative to the starting point.

Results

Univariate analyses of variance were performed on all of the measures collected and Table 1 presents a summary of these results. Several aspects of the data are worth elaborating. First, every measure indicated a significant pilot effect which for the most part was due to pilot by motion cueing and pilot by system delay interactions. Two pilots performed best with the longer system delay and occasionally a pilot did not show the typical ordering of performance under the conditions of motion cueing. Second, both the conditions of motion cueing and system delay showed significant effects on most of the measures. In general better performance was associated with the shorter delay and the ordering of the motion conditions was for best performance to be under the MB condition and the poorest under the FB condition with the G condition inbetween. Results from the vector combination of the longitudinal, lateral and altitude errors are presented in Figure 1. The pattern of these data is typical for most of the measures we took, and of some interest is the lack of a motion cueing by system delay interaction. Some preliminary data indicated that such a result may occur, but our formal experiment showed no evidence for it. Third, little effect is produced by motion of the ship model. And last, interesting interactions of ship motion by motion cueing were seen in the altitude error and collective input. Most of the other interactions of the main variables of the experiment are due to pilot differences.

Discussion

The aim of this study was to determine the effects a motion platform or G-seat would have on helicopter hover, and clearly both produced improved performance over that associated with a fixed-base simulation. A depressing result was the tendency for pilots to show idiosyncratic patterns of performance under these conditions. For the most part, best performance was associated with the moving base simulation, yet one individual performed best in the FB condition. The same can be said for the delay variable. In general, better performance will be seen with shorter delays, but again, two individuals performed better with the longer system delay. Perhaps, the least explainable result is the ship motion by motion cueing interaction seen with the rms measures of altitude error and collective activity. The greatest altitude error was generated with the moving ship model under the G condition while the lowest collective activity was seen under the

same condition with the ship motion off. We have no satisfying explanation for this result. The rest of the significant interactions involve the pilot variable and are due to variability in their manner of responding to the conditions we studied. The main effects of the conditions of motion cueing and the system delay are clear, with most measures showing the same pattern as in Figure 1, yet the lack of consistency on the part of the experienced pilots poses a problem for this sort of work.

Summary

This study examined the effects that various conditions of motion cueing, system delay and ship model motion had on helicopter hover performance. A within-subjects study of twelve experienced pilots indicated that while the main effects of the motion and delay variables were clear, pilot differences accounted for numerous higher-order interactions. Best performance was associated with motion base cueing with the short delay, followed by similar G-seat cueing, with poorest performance seen under the fixed-base condition. The longer delay uniformly elevated these performances, but little difficulty was added by moving the ship model. The implications these data have for helicopter simulation seem clear, yet a somewhat depressing result was the tendency for non-uniform effects to be seen across pilots.

References

Fortenbaugh, R.L. Mathematical models for the aircraft operational environment of DD-963 class ships. Vought Corporation Report 2-55800/8R-3500 September,1978.

McKissick, B.T., Ashworth, B.R., Parrish, R.V., and Martin, Jr., D.J. A study of the comparative effects of various means of motion cueing during a simulated compensatory tracking task. ACM/IEEE/SCS 13th Annual Simulation Symposium, March, 1980.

Parrish, R.V. Platform motion for fighter aircraft - Let's be realistic. AIAA Flight Simulation Technologies Conference, September, 1978.

Parrish, R.V., Houck, J.A., and Martin, Jr., D.J. Empirical comparison of a fixed-base and a moving-base simulation of a helicopter engaged in visually conducted slalom runs. NASA TN D-8424, Mary, 1977.

Puig, J.A., Harris, W.T., and Ricard, G.L. Motion in flight simulators: An annotated bibliography. NAVTRAEQUIPCEN IH-298, July, 1978.

Table 1. Results of Univariate Analyses of Variance

	System Error				Plant States		Control Inputs			
Source of Variance	Y	X	Alt	Vector	Aircraft roll	pitch	Cyclic pitch	roll	rudder	collective
Rep	**			**				**		
P	**	**	**	**	**	**	**	**	**	**
S						**			*	
PXS					**		**	**	**	**
D	**	**	**	**	**	**	**	**	**	**
PXD	**	**	**	**	**	**	**	**	**	**
SXD	**	**		**						*
PXSXD		*			*		**	**	**	**
M	**	**	**	**	**		**	**		
PXM	**	**	*	**	**	**	**	**	**	**
SXM			*							**
PXSXM	*	**		**	**	**	**	**		*
DXM							*			
PXDXM							**	**	**	**
SXDXM	*							**	*	
PXSXDXM		*	*		*	**	**	**	**	**

Rep = replications, P= pilots, S=ship motion, D=delay, and M=motion conditions. *=P $<$ .05, ** = P $<$ .01

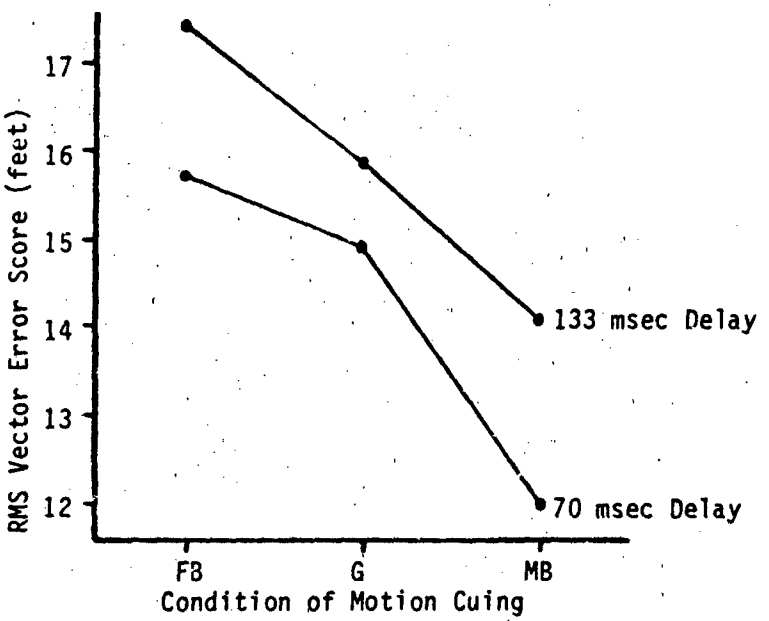


Figure 1. Vector Error Score Averaged Over Conditions of Ship Motion.

PILOT-COMPUTER INTERFACES FOR COMPLEX AERIAL NAVIGATION: SIMULATIONS USING A MICROPROCESSOR

Dennis B. Beringer

University of Illinois at Urbana-Champaign

Three configurations of a computer-generated 4-D aerial navigation system were evaluated using a PLATO-based digital flight simulation combining local and central processing. Each of 24 experienced pilots received training using either a keyboard entry/static map, keyboard entry/dynamic map, or touch entry/dynamic map system. During area navigation courses pilots exercised continuous flight control, updated navigation information, entered digital data, and plotted amended courses. The touch-map reduced training and task execution times and virtually eliminated errors on the plotting task. Flight simulation evaluations showed that the touch-map reduced tracking errors while increasing secondary task performance and course plotting accuracy. It was concluded that a touch entry system could significantly reduce cockpit workload.

INTRODUCTION

Problem

The National Airspace System is currently under examination for possible revisions to its structure and method of operation. New demands are being placed on the system in the areas of capacity and operational limits, forcing a reevaluation of airborne flight control (pilot and aircraft) and ground flight control (air traffic controller and computer). This examination extends to the subsystems contained in each larger system. The current study examined the possibility of transitioning from present 3-D flight restrictions with "informal" time scheduling to rigid "metering and spacing" techniques dependent upon time-phased (4-D) clearances. This would allow rigid control of traffic and eliminate conflicts excepting those stemming from emergencies.

The most enduring system configuration solutions seem not to reside in the modification of present systems, but in the development of new ones (Beringer, 1979). A continuing problem in the development of new man/computer systems is the establishment of efficient and natural communication between the human and machine elements. This aspect of system design takes on increasing importance with advances in computer technology (Carel, McGrath, Hershberger, and Herman, 1974). Fenwick (1970) suggested that the power of the computer could be compromised significantly if interface designs and data input/output strategies failed to provide a common frame of reference for the operator and computer. It is apparent that design of the operator/computer interface requires serious attention if the system, as a whole, is to function at or near its potential.

PRECEDING PAGE BLANK-NOT FILMED

Design Solution

Three characteristics were considered highly desirable for the new system configuration. First, it was important that the cockpit displays be integrated and provide maximum usable information with minimum clutter. Second, the operational procedures should be simple and highly structured so that minimum memorization was required. Finally, it was desirable that in dealing with map-related information both the pilot and computer be able to "draw" on the same display surface (Licklider, 1960). The final configuration developed satisfied these requirements, combining the data entry device and electronic map into a single control/display unit (CDU). An infrared touch panel overlayed the plasma-panel display surface, providing 256 touch points by which the pilot could access or change data through either software-generated function keys or by direct manual designation on the map. A vertical situation display was included as well as a message area displaying procedural instructions and course amendment information (Figure 1).

METHOD

Three system configurations were chosen for evaluation, including the touch-map and two keyboard-based systems (static and dynamic map configurations). The keyboard-based systems represented currently deployed systems modified for the 4-D environment and used three display areas; map, flight instruments, and nagivation CDU. The touch-map used a single display as the map was also the CDU. The PLATO system (Bitzer & Johnson, 1971) was selected as the simulation base. Three terminals connected to the central time-shared system were used, each having a plasma dot-matrix display screen with 512-by-512 resolution. Flight instrumentation was generated on a terminal containing an 8080 microprocessor, providing the means for real-time analog control inputs (stick & throttle) and flight instrument indications. All data collection, reduction, analysis, and graphics were done on the same system, requiring no data transfers or storage medium changes.

Experimental Tasks

During normal area navigation (RNAV) the pilot is required to perform two tasks; continuous flight control and stored navigation information updating (selecting the next waypoint to fly to). Two additional tasks were selected to exercise fully each of the systems. A digital data entry task was included to simulate the situation in which new waypoint coordinates are trasmitted to the pilot by air traffic control (ATC) and must be entered for nagivation. The second was the plotting of a new waypoint location within boundaries set by ATC that allowed some discretion to be exercised by the pilot (placing the waypoint in a defined area). In addition to normal operating tasks and one of the data entry tasks, a self-paced digit-cancelling side task was employed (Beringer, Williges, & Roscoe, 1975). This served as both a loading factor and a measure of effort being expended on the primary task.

Experimental Design

It was desirable that several environmental and task variables likely to affect system performance be examined across ranges that would adequately represent those expected in the operational environment. A central-composite design (CCD) was selected for this purpose (Myers, 1971) and was used to generate most of the experimental conditions. Five independent variables likely to influence performance (vertical path gradient, degrees of course change, required groundspeed, across-course drift, and along-course drift) were examined at five levels each. The design combined a 2^5 fractional factorial (half replicate) with six centerpoints and 10 axial points (extremes of the five levels) (within-subject) and was replicated for each system configuration.

An auxiliary mixed factorial design was used to evaluate performance on the waypoint plotting task, combining three levels of course profile difficulty (within-subject) with system configurations (between-subjects). The low- and medium-difficulty conditions were similar to those appearing in the CCD while the high-difficulty condition combined extreme values of all five independent variables (the most difficult condition in the experiment). The plotting task was presented on only three trials due to its small probability of occurance relative to the digital data entry task.

Subjects and Procedures

Subjects were 24 experienced instrument pilots recruited from the local community. Each received a preflight briefing booklet to study and a pilot experience questionaire to complete prior to the first training session. At the beginning of the first session each pilot was tested on three component tasks; digit cancellation, keyboard data entry, and 4-D navigation tracking. These baseline measures, along with questionaire measures, were used to assign pilots to groups such that each group contained eight pilots by the end of the experiment. Post-hoc comparisons indicated no reliable differences in group composition produced by matching. The session then proceeded with initial entry procedures training (to criterion) followed by inflight training. Training continued over several sessions until performance stabilized (generally 6 to 10 hours with 12 hours set as the upper limit). No session was allowed to exceed four hours.

During the final four-hour session each subject experienced the 32 conditions of the CCD in one of eight serially balanced orders. These conditions were arranged to provide the appearance of a continuous flight with the exception that brief straight-and-level segments of "hands-off" flight were interposed to minimize effects from the preceding condition. Pilots took hourly breaks of from 5 to 10 minutes throughout data collection. After a brief break at the conclusion of the CCD trials each subject flew the conditions described by the auxiliary factorial in order of increasing difficulty. When these trials had been completed the subject was debriefed and given a demonstration of the alternate system configurations.

FINDINGS

Training

Although digital data entry procedure training results were comparable across all three groups, waypoint plotting results showed that the touch-map required fewer trials to criterion, produced fewer errors, and facilitated more rapid task performance at exit. The static map group (Group 1) required an average 10.75 trials to criterion, averaged 4.62 errors, and had an average task completion time, at exit, of 69.64 seconds. The dynamic map group (Group 2) averaged 8.37 trials to criterion, 3.0 errors, and a task completion time of 50.46 seconds. The touch-map group (Group 3) averaged 7.75 trials to criterion, .25 errors (errors in digital entry, not waypoint location), and a task completion time of 32.96 seconds. Scheffe post-hoc comparisons showed that groups 2 and 3 differed reliably from 1 on trials to criterion ($p < .05$), 3 differed reliably from 1 and 2 on errors ($p < .005$), and each group was reliably different from every other group on task completion times ($p < .005$).

Analysis of initial inflight training indicated a reliable difference in performance between the keyboard groups and the touch-map group for waypoint plotting. Group 1 averaged 4.25 trials to criterion, Group 2 4.87 trials, and Group 3 1.38 trials. Post-hoc comparisons indicated that groups 1 and 2 did not differ reliably but when combined differed reliably from Group 3 ($p = .01$). These findings suggest that the touch-map with an embedded keyboard can adequately perform digital data entry tasks while accelerating training and improving performance on plotting procedures.

CCD

Data for eight replications of three complete central-composite designs was collected and analyzed using regression techniques. This produced second-order multivariate regression equations for both variable error measures (lateral, vertical, and time RMSE; adjusted side-task rate) and constant error measures (mean lateral, vertical, and time errors). Twenty-four main effects were reliable for one or more groups. Fifteen were in the expected direction (positively correlated with independent variables) and showed group differences. In 11 cases Group 3 exhibited consistently better performance, in 7 cases Group 3 exhibited better performance at the high end of the range, and in the remaining cases some unexpected cubic effects were present. Similar patterns appeared in the 29 reliable interactions.

Auxiliary Factorial

Two-way analyses of variance showed reliable group differences for lateral RMSE and adjusted side-task processing, favoring Group 3 in each case. Group 3 had the smallest lateral RMSE (3.01) and smallest side-task drop (-.049 bits/second), followed by Group 2 (4.86 and -.283 respectively) and Group 1 (3.63 and -.390 respectively). Both effects were reliable at $p < .05$. Group 3 also achieved the only above-baseline side-task performance (+.219 in the easiest condition). Group 3 produced plotted solutions closer to the center of the designated area than groups 2 or 1 (.411, .632, and .968 miles, respectively).

All Group 3 solutions to the medium-difficulty problem were identical, however, as the location fell directly on a touch entry point. The expected ANOVA could not be performed but analysis at the high-difficulty point showed a near-reliable effect favoring the touch-map (.42 miles) over groups 2 (.64 miles) and 1 (1.046 miles) ($F(2,21) = 3.084$, $p = .067$). There is little doubt that the overall effect favors the touch-map for waypoint positioning accuracy.

Simulation System Reliability

Three areas of failure were encountered during operation of the hybrid simulation system; central system, communications, and local hardware. Central system failures were most frequent, affecting 48% of the 75 sessions (300 hours). Of these 36 delays, three forced termination of flights. The remainder averaged .3 hours in length (range = .1 to 1.9). Communications errors between the 8080 microprocessor and central system were second, affecting 24% of the sessions with an average 3.4 errors per affected session (range = 1 to 13). This resulted as a combination of slow local-to-central transmission rates (4 8-bit bytes/second) and processing delays in the central system (despite safeguards in local and central software). Local hardware failures affected only 5% of the sessions and only one failure involved the intelligent (8080) terminal. These results favor a dedicated micro- or mini-computer simulation as two major sources of failure could be eliminated.

SUMMARY

The inclusion of an interface device allowing direct manual access to navigation data in map form can be beneficial. The touch-map system appeared to reduce required operating time by simplifying procedures and providing a direct, intuitive link between the pilot and nagivation computer. This allowed increased effort to be directed towards other tasks, effecting a reduction in cockpit workload. The central-composite design provided an economic means for sampling the environmental space to find areas affected by variable manipulations. The fact that some group differences appeared only at extreme levels may aid in subsequent investigations by allowing their focus to be reduced to areas where an effect is present. Simulation system reliability results suggest that dedicated systems, within certain limits, should be considered over time-shared or hybrid systems when the simulation is computationally and graphically within the capabilities of a microprocessor system.

REFERENCES

Beringer, D. B. The design and evaluation of complex systems: Application to a man-machine interface for aerial navigation. In C. K. Bensel (Ed.) Proceedings of the twenty-third annual meeting of the Human Factors Society. Santa Monica, CA: Human Factors Society, October 1979, 75 79.

Beringer, D. B. Williges, R. C., and Roscoe, S. N. The transition of experienced pilots to a frequency-separated aircraft attitude display.

Human Factors, 1975, 17(4), 401-414.

Bitzer, D. L. and Johnson, R. C. PLATO--A computer-based system used in the engineering of education. IEEE Proceedings (Special Issue on Engineering Education), 1971, 59, 960 - 969.

Carel, W. L., McGrath, J. J., Hershberger, M. L., and Herman, J. A. Design criteria for airborne map displays. Culver City, CA.: Hughes Aircraft Co., Final Report C2151-003, 1974.

Fenwick, C. A. The pilot interface in area navigation. Cedar Rapids, IA.: Collins Radio Company, paper prepared for the Human Factors Society 14th Annual Meeting, 1970.

Licklider, J. C. R. Man-computer symbiosis. IEEE Transactions on Human Factors in Electronics, 1960, HFE-1(1), 4-11.

Myers, R. H. Response suface methodology. Boston: Allyn & Bacon, 1971.

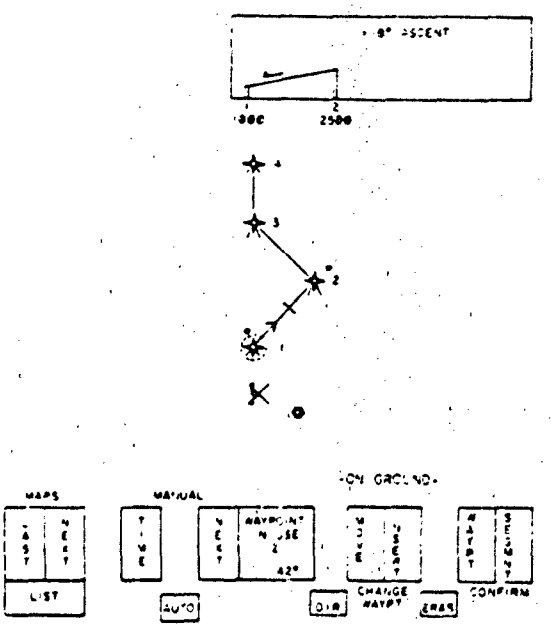


Figure 1. Integrated touch access control/display unit.

How Should You Talk to a Semi-Intelligent Machine?

James K. Arima
Anthony G. Quartano
Naval Postgraduate School
Monterey, California 93940

Abstract

15 P-3C and 15 non-P-3C aviators were presented several situations on a simulated P-3C pilot's display and required to verbalize commands to obtain the final display in a particular set--e.g., "center aircraft." The latency, syntax, and number of different commands differed for the two groups, but the content and use of two-word commands were similar. The results suggested that the vocabulary for voice-actuated systems being substituted for existing manual systems should conform to the latter. The vocabulary for completely new systems could use one- or two-word commands (with the syntax being immaterial for two-word, sequential commands) and still conform to the natural verbal habits of aviators.

The possibility of communicating with machines in natural, spoken language is now so imminent (Robinson, 1979a, b) that many persons are contemplating the benefits and problems that might occur in the interface of mankind with such intelligent machines. Should they be able to provide information processing capabilities parallel to the existing audiovisual-manual channel, the benefits would be immediate and tremendous. If, as it has so often been shown, the new capabilities confirm that man is a single-channel processor, then much research will be required to determine the trade-offs that will enhance human performance when the man-machine interface includes spoken language. Even if the total information processing rate remains constant, there may be many situations in which it would be advantageous to use the voice channel rather than, for example, manually keyed instructions (Beek, et al., 1978; Naval Training Equipment Center, 1979).

Because utilizing voiced commands to initiate unique, discontinuous events is much simpler than natural-language communication in a continuing activity, attempts have been made to apply currently available techniques to operational systems that fit this framework. This case only requires the machine to recognize isolated commands for a finite number of operations. Typically, the machine is trained to recognize the commands of a particular individual. Thus, these devices are often called learning machines in contradistinction to teaching machines. For them, the smaller the size of the vocabulary, the more phonetically distinct the individual words, and the more consistent the speaker, the greater will the efficiency of the machine be. Under these conditions, the capability of the device can be greatly enhanced without a degradation in performance if sequential, branching combinations of words are used. For example, if 1 of 4 choices is made in the first command and 1 of 4 additional choices is made in the second command, 16 operations could be differentiated. Obviously, this situation is much more simple for the machine than learning to discriminate 16 different commands at one time.

added in three displays following the "before" display, (3) the "before" slide labelled the aircraft symbol as being in the wrong location and the "after" slide showed it offset a considerable distance, (4) the "before" display showed the aircraft symbol at the edge of the display and the "after" display showed and identified it as being at the exact center of the display, and (5) a set of three slides showed the aircraft and sonobuoy symbols getting closer together while the display scale indicator increased. Short titles were given these sets for identification as follows: (1) position, (2) track, (3) correct, (4) center, and (5) scale.

Each subject was run individually in a sound-deadened, enclosed booth. Displays were back-projected on a milk-glass screen from outside the booth. The subject was given a recorded orientation that described the display system, experimental task, and procedures. A sample set of "before" and "after" displays showing aircraft altitude was included in the orientation. The subject was then presented 3 blocks of the 5 sets. The sets were in random order within each block. Each display remained visible for 10 sec., and there was a 10-sec. dead time between sets. Response time to the "after" display was determined using a stopwatch, and the subject's response was electronically recorded.

Results. An analysis of variance of response times to the "after" display showed significant effects ($p < .001$) for learning (blocks), groups, and tasks. None of the interactions was significant. The mean response times (in sec.) and their standard deviations were for all subjects: 5.43 (2.91), P-3C: 4.23 (1.80), and Other: 6.62 (3.32). The mean response times to Scale (3.98) and Track (4.54) were different from each other and quicker than the other three sets, which were all at approximately 6 sec. The qualitative results are summarized in Table 1, which shows the number of different commands used, the number of one-word commands, and the percent of subjects using the display term or its derivative in their commands.

Table 1. Number of Different Commands and One-Word Commands (in parentheses) by Display, Group, and All Subjects, and Percent of Responses Using Display Term.

Display	P-3C	Other	All	Percent
Position	5 (2)	10 (4)	11 (4)	93
Track	4 (1)	17 (4)	17 (4)	88
Correct	11 (0)	22 (7)	26 (7)	49
Center	6 (1)	10 (2)	11 (2)	93
Scale	5 (0)	18 (2)	18 (2)	89
Total	31 (4)	77 (19)	83 (19)	82

The table shows that the P-3C pilots used many fewer commands and very few one-word commands. The amount of commonality present in the commands is revealed by the fact that the total number (all) of commands for most displays was essentially the same as that for the Other group of pilots. That is, the Other set essentially included the commands used by the P-3C group. The number of one-word commands differs significantly for the two groups. The response initiation times, however, did not differ within each

In the foregoing example, it would be much simpler for the human operator to choose 1 out of 16 choices. In such instances, when an attempt is made to stretch the available technology while maintaining high performance standards, the human partner in the system may be required to pay the costs. An arbitrarily limited set of words chosen to maximize phonetic dissimilarity and connected together in rigid sequences places a learning burden on the operator and greatly increases the chances for error under operational and, particularly, stressful conditions. Accordingly, it would be wise to determine the verbal habits of the operator that could be used to make the language natural, easy to learn, and less error prone. Adams (1975) and his associates have demonstrated that meaningfulness and natural language mediators facilitate the learning of word pairs. On the other hand, Jung (1966) has found that word associations may be quite situation specific, when they are presumed to reflect the underlying verbal habits of the individual. Thus, in the context of a learning-machine vocabulary, meaningful and natural language should facilitate the learning and application of the vocabulary but the specific situation and the individual's past experience in that situation should also influence the appropriateness of words and their combinations.

In view of these various factors that could influence the choice of a vocabulary for a particular application, this study examined the role of situation-specific experience on the free choice of words for a command language and the degree of verbal commonality that exists in a specific situation among a relatively homogeneous group of operators with different backgrounds.

Method. The general approach was to have experienced military pilots free associate instructions to patterns on a simulated tactical display. One group of pilots was quite familiar with the particular display, which they normally controlled through a keyset panel. Another group of pilots had no experience with the particular display but had experience with electronic displays in combat aircraft. The first group was representative of a situation in which voiced controls were directly substituted for manual controls. The second group represented individuals who were newly assigned to a voice-actuated system. The subjects were 30 volunteer, male students at the Naval Postgraduate School with over 1,000 flight hours in military aircraft. Half of them had experience in P-3C antisubmarine warfare (ASW) aircraft and will henceforth be referred to as the P-3C group. The remaining 15 pilots will make up the "Other" group.

The task presented to the subjects displayed a situation labelled "before" and another labelled "after" on a highly simplified, simulated P-3C display. Two tasks had intervening situations that were unlabelled. Upon presentation of the "after" display, the subject voiced a command that would have brought about the "after" display. To provide a perspective of scale and motion, all displays had a sonobuoy symbol and a symbol of the aircraft. The item that was changing or changed was indicated by an arrow. A verbal flag was used in two of the sets to ensure that the situation was unambiguous. All display changes represented operations that were attainable with a single keyset in the P-3C aircraft. In all, there were five sets of displays: (1) The "after" display showed the latitude and longitude of the aircraft, (2) symbols showing the track of the aircraft were

group for the one- or two-word commands. In spite of the number of different commands used, the proportion of commands that used a common reference word (the display identifier) was over 80 percent for all subjects and displays. The two-word commands were classified as imperative (verb first) or nominative (noun first). Examples would be "enlarge scale" versus "scale enlarge." The two types were significantly different for the P-3C group with more than 60 percent of the responses being of the nominative type. There was no significant difference for the Other pilots.

Summary and Conclusions. The P-3C pilots showed their experience with the specific situation in quicker reaction times, greater commonality in commands, and preference for two-word commands in the nominative order. The latter preferences reflect the instructions on the P-3C keyset panel. While there was a greater variety of commands for the Other pilots, there was considerable commonality with the P-3C pilots in the use of key terms and two-word commands. They were different in their lack of preference for syntactical word order in their two-word commands and while the total number was not great, they gave more different one-word commands. All pilots initiated two-word commands as quickly as one-word commands. These findings suggest that substitution of voice systems for an existing manual system should use the familiar, existing terms and that they would quickly be acquired by individuals new to the system. For entirely new systems, two-word commands should be used as a rule and when discrimination of commands would be helped by longer terms. When reaction time of the system is important, one-word commands should be used. When two-word commands must be used in sequence in a new system with a relatively large number of actions, it does not seem to matter whether the item of equipment or the function comes first. That is, "Fire Sparrow" and "Sparrow Fire" are equally natural to a pilot naive to the system. But the system should be consistent in the syntax used for all functions. The trade-off in word order can be made on the basis of engineering convenience. Finally, the limitations of this study should be realized in the artificiality of the task situation, and it would be wise to run a similar experiment for decision making in a real-world, system design context.

References

Adams, J. A. Learning and memory: an introduction. Homewood, Illinois: Dorsey, 1975.

Beck, B., Neuberg, E. P., & Hodge, D.C. An assessment of the technology of automatic speech recognition for military applications (Technical Memorandum 5-78). Aberdeen Proving Ground, Maryland: U.S. Army Human Engineering Laboratory, March 1978.

Jung, J. Experimental studies of factors affecting word associations. Psychological Bulletin, 1966, 66, 125-133.

Naval Training Equipment Center. Minutes: voice technology for systems application. Orlando, Florida: Author, March 1978.

Robinson, A. L. Research News: More people are talking to computers as speech recognition enters the real world. Science, 1979, 203, 634-638. (a)

Robinson, A. L. Research news: Communicating with computers by voice. Science, 1979, 203, 734-736.

Use of Imagery-Based Training Materials for Learning Procedural Tasks

Gary G. Miller
Technical Training Division
Air Force Human Resources Laboratory

An exploratory study was conducted to determine the effectiveness of an imagery-based training strategy for learning procedural tasks involved in the operation and maintenance of F-111D automatic test equipment. A total of 53 technical training students were divided in an imagery training group and a control group and were trained on four procedural tasks followed by a delayed recall test on Task 1. Results indicated a significant difference on Task 1 favoring the imagery group and no differences on delayed recall. The control group scored significantly higher on a short verbally oriented task. Interference and practice effects were offered as explanations for the pattern of results. Future experiments varying task difficulty and controls for interference are required to derive more conclusive results.

Imagery as a Training Technique

Interest in the potential of imagery as a learning strategy has gained momentum in the past few years, partly as a result of the growth and influence of cognitive psychology. A substantial body of literature has accumulated that shows imagery to be a powerful strategy for learning verbal materials (Paivio, 1971). Most of this research has been concerned with the effects of imagery on the learning of verbal paired-associate tasks and, to some extent, narrative verbal materials. However, very little is known about the application of imagery strategies to complex sequential-learning tasks or nonverbal tasks, both of which are typical of those to be learned in the operation and maintenance of military systems.

Two recent studies which have used imagery strategies for learning procedural tasks have been reported by Prather (1973) and Johnson (1978). Prather used mental practice to encourage mental imagery for learning T-37 landing patterns and found that mental practice combined with actual practice was more effective than actual practice alone. Johnson investigated imagery strategies for learning conveyer-line production operation and found that strategies requiring the trainee to provide his own cueing and feedback from memory are effective for increasing retention in procedure-following skills.

The purpose of the study was to evaluate the feasibility and training-effectiveness of an imagery technique in training tasks

involving maintenance of avionics components from the F-111D. The technique purports to use visual imagery to enhance learning.

The major question posed in this research was whether the imagery technique produces learning of procedural tasks, as measured by delayed recall, that is equal to or superior to a control method that would appear to lack some of the imagery-evoking features of the imagery technique.

Method

The 6883 Test Station, for maintenance of avionics components of the F-111D aircraft, was selected as the test vehicle for the imagery evaluation. Experimental and control groups of Air Force students were presented with either imagery-based (experimental) or verbally oriented (control) training materials, which they used for studying four task sequences on the 6883. Both groups were tested for learning of the steps comprising each task. A retention test was also given for task 1. Several groups were run under either the experimental or control condition over a five-day period.

Subjects

Fifty-six Air Force students in training at Lowry AFB, CO were used as experimental and control subjects. They participated in the experiment while awaiting their next block of technical training. All subjects were familiar with electronic equipment but naive regarding the 6883 Test Station. Subjects were run in groups of 7, with all subjects in a group participating in the same treatment condition. Subjects were randomly assigned to groups over a five-day period.

Materials

The imagery training material used by the experimental group consisted of a study guide and a practice environment. The study guide presented the four tasks in a workbook format, using graphics to translate verbal instructions for each step into a highly visual depiction of the required physical movements. The practice environment was a two-dimensional photographic model of the 6883 Test Station, accompanied by line drawings of each section that were specifically related to tasks, could be slipped over the model, and had answer keys underlying them.

The conventional training materials were in the form of a booklet, and presented the same four tasks in the way that they appear in the technical orders. The control group was also allowed to use the photographic model of the test station, but without the accompanying experimental materials.

Procedure

Subjects were given a brief introduction to the experiment and to the training materials they would use. The experimental group was given a more detailed orientation, since they had had no prior experience with materials of this type. The control group would practice tasks only by means of mental review, while the experimental group would make the appropriate marks on the practice environment as they followed the study guide. As a measure of their imaging ability, subjects were also given a preliminary memory test for geomentric shapes.

Each group was put through four cycles of training and testing (one for each task). To perform the learning tests, subjects wrote down the procedural sequence of the task in as much detail as possible. After completing the last cycle, subjects took a retention test for the first task.

Finally, subjects were asked to fill out an attitude questionnaire in which they gave a subjective assessment of the training materials they had used.

Results

Table 1 shows the mean percentage correct for groups on the immediate and delayed recall tests of Task 1. These data were analyzed by a 2 x 2 repeated measure analyses of variance. The

Table 1

Mean Percentage Correct for Groups on the Immediate and Delayed Recall Tests for Task 1

	Test	
Group	Immediate	Delayed
Experimental, N=26	64.8	51.7
Control, N-27	52.6	51.0

factors were experimental vs. control group and immediate vs delayed recall testing. The interaction of groups X test was significant ($F(1,51) = 14.61$, $p < .01$). The experimental group scored significantly higher on the immediate recall test for task 1, but no differences were found on the delayed recall test. The main effect for tests was also significant ($F(1,151) = 22.78$, $p < .01$) indicating that there was a significant difference on scores between the immediate and delayed recall tests. The analyses of variance table is presented in Table 2. Table 3 shows the mean percentage correct for groups on the learning tests of tasks, 2-4. These data were

analyzed separately for each task by a t-test between the experimental and control groups. As can be seen in Table 3, the

Table 2

Summary Table of the Analyses of Variance on the Immediate and Delayed Recall Test Scores of Task 1

Source	SS	df	ms	F
Between				
A (Imagery)	1109.68	1	1109.68	2.19
Subjects within Groups	25837.58	51	506.62	
Within				
B (Delayed Recall)	1398.34	1	1398.34	22.78*
A X B Interaction	896.80	1	896.80	14.61*
Subject within Groups X B	3131.36	51	61.38	
Total	32373.76	105		

* $p < .01$

differences between the experimental and control groups were not significant on tasks 2 and 4. The difference on task 3 was significant, with the control group scoring higher than the experimental group ($t(51) = 3.49$, $p < .002$).

Table 3

Mean % Correct for Groups and t scores on the Immediate Recall Tests of Tasks 2 - 4

Source	Task 2	Task 3	Task 4
Experimental	85.6	83.7	62.4
Control	79.3	95.9	67.0
t-score	1.30	-3.49*	-.89

*p .002

Discussion

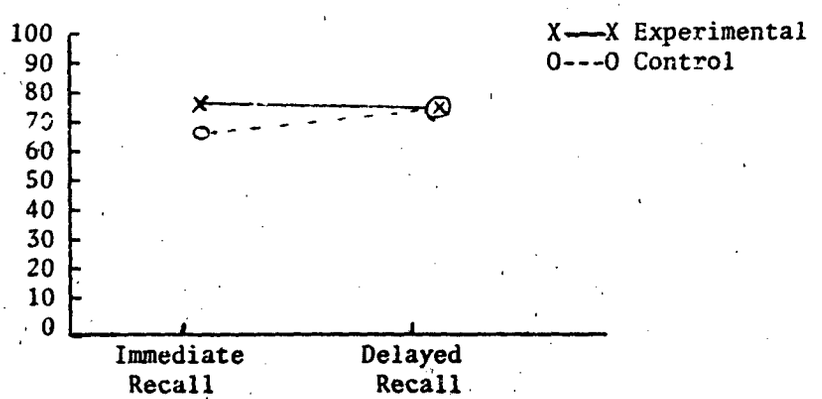
In general, these evaluation results were mixed with respect to the effectiveness imagery vs. conventional training materials. Three of five recall tests revealed no significant differences between the experimental and control groups; one test resulted in significantly higher performance by the control group; and one test resulted in significantly higher performance by the experimental group. The latter significant difference favoring the experimental group on task 1 is considered to be important since the primary analysis

performed in this study was between performance on task 1 and delayed recall testing on task 1. A number of issues are suggested by these results.

The interaction found on the learning and retention tests of task 1 is shown in Figure 1. As can be seen the experimental group demonstrated superior knowledge of the task when tested immediately after the study period. It is reasonable to conclude that some aspect of the imagery training materials was the cause of that superior performance. It can also be seen that this advantage had diminished by the time of a retention test for the task, administered by the time of a retention test for the task, administered approximately two hours later. This result is somewhat unexpected, since the amount of forgetting from initial learning test to delayed recall testing should be the same for both groups. The level of forgetting experienced by the experimental group is not large in view of the potential interference caused by performance on tasks two through four. However, the constant performance level recorded by the control group may indicate that the performance of tasks two through four by the group was facilitating rather than interfering.

Figure 1

Mean Percentage Correct for Groups on the Immediate and Delayed Recall Tests for Task 1



One possible explanation for the result is that the use of a visual imagery strategy in which many of the same controls and indicators had to be visualized in different positions from task to task may have been interfering. In addition, the experimental group was trained on a graphic representation of the test equipment and was required to transition to a photographic mock-up for testing. The control group, however, used the same materials for learning as well as testing. As a result, the experimental procedures may have inadvertently contributed to a practice effect for the control group which produced almost no loss from learning to delayed recall testing several hours later.

The success of the experimental group on task 1, in light of the equal performance results obtained on the recall test, makes this study inconclusive as to the effectiveness and value of the technique. As an exploratory evaluation, however, these results do suggest the need for refinements to both the technique and the evaluation method. It is recommended that further work control for possible interference effects and manipulate task difficulty in order to determine the technique's relative effectiveness for highly complex situations. Also, the various elements of the imagery technique should be isolated in order to determine what element or combination of elements is sufficient for effective training.

References

Paivio, A. Imagery and Verbal Processes. New York: Holt, Rinehart and Winston, 1971.

Prather, D. Prompted mental practice as a flight simulator. Journal of Applied Psychology, 1973, 57(3), 353-355.

Johnson, S. L. Retention and transfer of training following a procedural task: Interaction of training strategy and cognitive style. (CALSPAN Report DJ-6032-M-1) January 1978.

Non-Conforming Behavior in the Early Volunteer Army

Larry J. Crisler
Millikin University

Abstract

A questionnaire asking subjects a number of fixed choice questions regarding their military status, conceptions of non-conformity, and a number of social psychological dimensions was mailed to 1000 soldiers stationed at an Army base in the United States during the early stages of the volunteer army experiment (VOLAR). Of the 312 respondents, all admitted to having engaged in at least one violation of the Uniform Code of Military Justice (UCMJ). There was no statistically significant relationship between rank and either the frequency or nature of such violations. However, nearly 32% of the total variance of the incidence of non-conforming behavior is accounted for by the respondent's age and concerns over (1) adequate career progress, (2) establishing satisfactory peer relationships, and (3) appearing appropriately masculine. These findings suggest techniques that might be employed to reduce such non-conforming behavior.

The very foundation of military organization is, and has traditionally been, discipline, yet relatively little is known about its failures--those instances in which military regulations have been evaded. Despite its implications for military leadership decisions, no universally accepted definition of non-conforming behavior can be found.

In this study, non-conforming behavior is a property that is conferred upon a situation, event, or act by others rather than an inherent characteristic of the situation, event, or act. Such a definition is consistent with the labeling perspective (Becker, 1963; Erikson, 1966), but it is not meant to preclude the possibility that some behaviors are in fact mala in se. The problem of who or what shall confer the property of non-conformity is greatly simplified in the military institution because it is authorized to exercise very comprehensive control over its members' lives. Within this paper, non-conforming behavior will be restricted to those actions which are characteristically singled out by the military judicial system as being counterproductive to the military mission. Those acts prohibited by the Uniform Code of Military Justice (UCMJ) presumably constitute the core activities that are considered to be detrimental to the military mission.

Taking the above as an operational definition of non-conforming behavior, there appear to be four general themes by which existing literature may be classified: (1) comprehensive studies in which non-conformity is largely a peripheral concern (Stouffer, 1949; Ginzberg, 1959; Moskos, 1971); (2) rank-specific works (Bidwell, 1961; Uyecki, 1960; Janowitz, 1971); (3) technical and contractual reports (Fisher, 1972a, 1972b; Caylor, 1969); and (4) personal statements and psychiatric analyses (Stapp, 1970; Sherrill, 1970). The following study is an attempt to objectively analyze non-conforming behavior within all ranks of the American military.

Method

Questionnaires were mailed in February of 1975 to 1000 randomly selec-

ted personnel assigned to an Army base within the Continental United States. No follow-up mailings were permitted. Anonymous replies were received from 312 subjects for a response rate of 31%. The respondents were very representative of the total U. S. Army in such respects as racial composition, age, marital status, and education (Department of Defense statistics 1975, 1976). There was a distinct response bias toward the higher ranking soldier; however, as NCO's and officers have rarely been included in previous studies of non-conformity, this would not seem to constitute a serious methodological problem.

The questionnaire was designed to be comprehensible to all military personnel and consisted of questions measuring the subjects' age, race, marital status, education, civilian social status, unit assignment, rank, military occupational specialty (MOS), present residence, and length of time in the service. The bulk of the questionnaire was composed of scales measuring (1) the subject's career expectations (2) anxiety concerning the progress of his career (3) anxiety concerning his relationship to other soldiers; and (4) the extent to which he recalled being prone to engage in aggressive acts prior to entering the Army. Finally, the respondent was presented with a list of twenty-nine violations of the Articles of the UCMJ for which he was asked to describe the seriousness of the violation, the risk involved in committing the act, and how many times he had engaged in the act.

Results

The data were analyzed by means of the SPSS computer software system; utilizing the conservative assumption of ordinal data where appropriate. Most relationships were evaluated by Kendall's Tau statistic, the ordinal equivalent of Pearson's correlation coefficient for interval data (Hawkes, 1971).

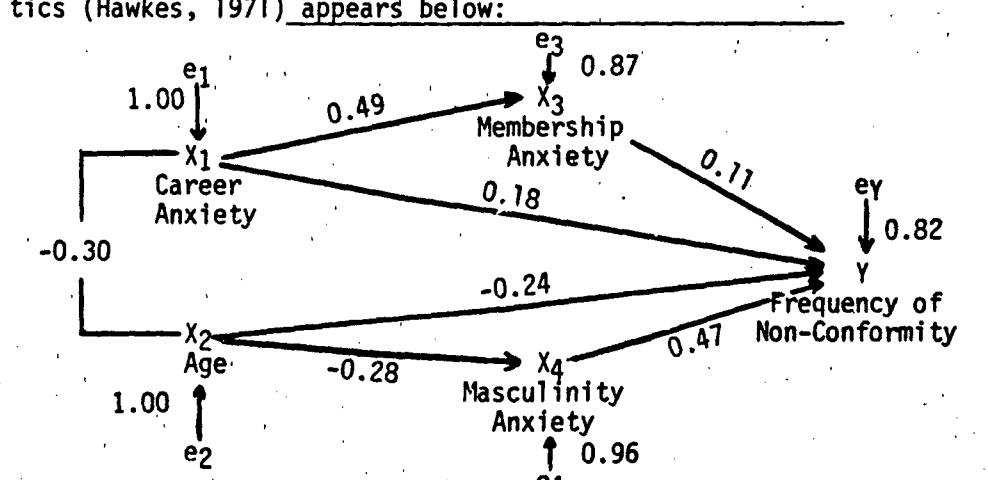
Each of the respondents reported committing at least one of the violations of the UCMJ. Among the most frequent violations were (1) acts that serve to enhance the violator's military career--warning a company of a forthcoming inspection that it might fare better, maintaining a duplicate set of immaculately clean equipment for inspections only, and altering scores on rifle qualification tests and (2) acts that even if undetected are liable to adversely effect the soldier's career--missing formations, drinking on duty, and smoking marijuana on duty. Rank had no statistically significant effect on either the frequency of non-conformity or the types of acts committed, although enlisted personnel are less likely to maintain a duplicate set of equipment and more likely to smoke marijuana on duty. When the 29 violations are ranked from most to least frequently committed for EM, NCO's and officers, the rank order correlations are consistently similar (rho between EM and NCO's is 0.84, between EM and officers it is 0.61, and between NCO's and officers it is 0.79--all are significant at $p<0.05$).

Previous research has suggested that non-conformity in the military is related to the soldier's age (Ginzberg, 1959; Fisher, 1972b), race (Ginzberg, 1959), marital status (Rose, 1951; Ginzberg, 1959), MOS (Fisher, 1972b), time in service (Stouffer, 1949), peer relations (Bialek and McNeil, 1968; Caylor and McFann, 1968), and education (Stouffer, 1949;

Uyecki, 1960). Our data provide moderate support for all but two of these frequently cited propositions. While these data confirm a statistically significant relationship between non-conforming behavior and race, they indicate that non-white soldiers are less non-conforming than white soldiers (χ^2=6.46, d. f.=2, p<0.05). These data also provide very strong support for a relationship between concern with establishing satisfactory peer relations and the incidence of non-conformity (Tau=-0.29, p<0.05).

It would seem that non-conformity in the military can best be explained not by the demographic characteristics of its members--neither those brought with them at entry nor those which result from military status, but rather by the soldier's concern over his relationships with others. This concern is reminiscent of Matza's (1964) theory of "drift" in which he proposed two central concepts--membership anxiety and masculinity anxiety--in the incidence of juvenile delinquency. We have already looked at the former. The latter--masculinity anxiety--is a very difficult concept to measure. The measurement problem is further complicated by the sex-segregated nature of the military. In attitude, military recruits seem to exhibit the traditional, stereotypical, characteristics of masculinity and might be universally expected to exhibit high levels of masculinity anxiety. There was, however, sufficient variance in the respondents' scores to permit statistical analysis employing this variable. As hypothesized by Matza, there is a strong inverse relationship between age and the degree of masculinity anxiety (Tau=-0.28, p< 0.05).

It is impossible to read the literature in military sociology without becoming sensitive to the concept of "career." Given the voluntary nature of current enlistments, the "public" nature of promotion lists, the policy of "riffing" (forcing into retirement those soldiers whose careers have fallen behind the "typical" pattern of promotions), and the rather open competition for advancement; anxiety over one's career seems likely to spill over into other areas. Career anxiety becomes the fourth central explanatory variable in the proposed explanation of self-reported non-conformity in the military. A causal diagram with ordinal regression statistics (Hawkes, 1971) appears below:



Multiple TAU=0.56, p<0.05 Multiple TAU^2=0.318

The overall causal logic of the argument generally follows Matza's (1964) perspective on juvenile delinquency. Such "objective" indices of the military environment as rank, MOS, years in the service, and the like are of far less importance in the understanding of military non-conformity than are the more socio-psychological concepts of anxiety--anxiety over career, peers, and machismo. Yet the source of such anxiety resides, ultimately, in the competitive demands of military life. That the proposed model accounts for nearly 32% of the observed variance signals a marked improvement over earlier explanations and serves to direct future research toward the analysis of competition in the military environment.

Discussion

At the risk of seeming overly pessimistic, we must admit to certain limitations that are apparent within this study. However, rather than detracting from the value of our findings, these limitations illustrate the course of future research. Practical considerations forced the use of a one-shot mailed questionnaire, and the resulting response rate of 31% constitutes the most obvious limitation of the research. Yet the strong associations obtained represent formidable evidence that the interpretations are substantively as well as statistically significant. Certainly the amount of non-conforming behavior reported for all ranks suggests that it is nearly universal within the American military institution and is not simply the result of sample bias. A second limitation of this study resulted from the use of cross-sectional data. Lacking information on the time-ordering of non-conformity and the antecedent variables--particularly the three anxiety variables--one is hard pressed to show beyond a reasonable doubt the direction of the several causal relationships. For example, does an individual engage in non-conformity because he is anxious about his career or does he become anxious because he has engaged in non-conformity and fears that he will be found out? The actual direction of the causal relationship would have serious theoretical manifestations. Only with longitudinal or time-series data could we determine the direction of these relationships in the fully satisfactory way.

We can speculate, in light of our findings, about changes that have been instituted to advantage with the American military and about changes that might be instituted in the future. Many overtly feminine tasks such as KP and routine cleaning and polishing have either been eliminated or reduced in the volunteer army. These changes should have reduced masculinity anxiety. Recreational facilities (gymnasiums, swimming pools, theatres) and gathering places (coffeehouses, libraries and PX's) should facilitate satisfactory peer relations. More actively assisting personnel who wish to marry or move off-post, providing monetary assistance for rent or moving dependents for all personnel and not just officers should also prove effective in reducing membership anxiety. The military career must, we suggest, be made less dependent upon narrowly prescribed protocol. More credit should be given for civilian acquired skills in the way of enlistment at ranks commensurate with duties. Often E-2's are found in positions that specify E-6's; should they not receive either the advanced rank or the increased pay?

Implementing changes of the sort suggested above has not called for, nor

will call for, drastic measures. While such changes will not eliminate non-conformity from the military institution they get at the core of what would seem to be, on the basis of our findings, focal concerns of a great many military personnel.

References

Becker, H. S. (1963) Outsiders. New York: Macmillan and Company.

Bialek, H. and M. McNeil (1968) Preliminary study of motivation and incentives in basic combat training. HumRRO Technical Report 68-6.

Bidwell, C. E. (1961) The young professional in the army. American Sociological Review, 26:360-372.

Caylor, J. S. (1969) Relationships between army recruit characteristics and first tour performances. HumRRO Technical Report 69-5.

Caylor, J. S. and H. H. McFann (1968) A follow-up study of the performance of army recruits in their first tour. HumRRO Paper 10-68.

Erikson, K. T. (1966) The wayward puritans. New York: Wiley and Sons.

Fisher, A. H. (1972a) Analyses of selected drug-related topics. HumRRO Technical Report 72-9.

Fisher, A. H. (1972b) Preliminary findings from the 1971 DoD survey of drug use. HumRRO Technical Report 72-8.

Ginzberg, E., et. al. (1959) The ineffective soldier, Volumes I-III. New York: Columbia University Press.

Hawkes, R. K. (1971) The multivariate analysis of ordinal measures. American Journal of Sociology, 76:909-926.

Janowitz, M. M. (1971) The professional soldier. Chicago: Free Press.

Matza, D. (1964) Delinquency and drift. New York: Free Press.

Moskos, C. C. (1971) Public opinion and the military establishment, Vol. 1. Beverly Hills, California: Sage Publications.

Moskos, C. C. (1970) The American enlisted man. New York: Russell Sage Foundation.

Rose, A. M. (1951) The social psychology of desertion from combat. American Sociological Review, 16:614-629.

Sherrill, R. (1970) Military justice is to justice as military music is to music. New York: Harper and Row.

Stapp, A. (1970) Up against the brass. New York: Simon and Schuster.

Stouffer, S. A., et. al. (1949) The American soldier. Princeton, New Jersey: Princeton University Press.

Uyecki, E. S. (1960) Draftee behavior in the cold war army. Social Problems, 8:151-158.

The Volunteer Army and Initial Adjustment Problems

E. R. Worthington
Brooke Army Medical Center
Fort Sam Houston, Texas

James Warren
University of Texas of Austin
Austin, Texas

Abstract

Several recent studies suggest the Volunteer Army concept is not working. Allegedly the poor quality of personnel obtained under this concept are unable to successfully adjust to military life. This study examines biographic, demographic and clinical data on 16 unsuccessful soldiers who were evaluated as unable to adjust and recommended for discharge. This data is compared to a similar study conducted in 1975. Demographic and biographic statistics of the 1979 and 1975 soldiers are not dissimilar. Psychiatric differences suggests the 1979 unsuccessful soldier is more alienated and maladaptive than his 1975 counterpart.

The question most often asked about the volunteer Army is, Does it work? Can the U. S. field a combat ready force of high quality soldiers? Recent data suggests that the Army is finding it difficult to procure quality soldiers in the quantity necessary to maintain a credible military deterent.

The U. S. Army Training and Doctrine Command (responsible for all soliers training) compiles statistics on the soldiers discharged during training for their inability to adjust or adapt to the military environment (Carney, 1979). During the first half of FY 1979 8.2 percent of male recruits were discharged compared to 6.4 percent for FY 1978. The Beard Study (Reed, 1978) finds that while thousands of young people are enlisting to build a dream most become dissatisfied and disillusioned. The study concludes that the Army is attracting more applicants from lower socioeconomic groups and with lower mental abilities. These people are not prepared to adapt to the rigorous demands of military life. A West Point instructor (Wesbrook, 1979) reports that the volunteer Army is comprised of alienated soldiers. The new enlistee brings into the Army a legacy of negative social values reflecting poverty, economic exploitation, poor education, discrimination, social instability and corruption. Wesbrook feels that the values and attitudes of today's volunteer soldier are incompatible with the requirements of sustaining an effective, disciplined, reactive Army.

This paper presents the results of a study which compared biographic, demographic, and clinical data of the 1979 unsuccessful male enlistee (N=16) with those of the unsuccessful male enlistee of 1975 (N=75). The unsuccessful soldier is one who, after evaluation by a mental health counselor, is seen as unable to adapt to the military environment and referred for immediate discharge.

Method

During the months of May through July 1979, 106 male enlisted medical trainees were evaluated by the Fort Sam Houston Community Mental Health Service. These soldiers were referred by self or their command because they were experiencing adjustment problems. Of this group, 16 were recom-

mended for immediate discharge from the Army due to their inability to adapt to the military environment. Data obtained by the intake counselor was then compared to a similar study conducted at an Infantry Training post in 1975 (Worthington, 1976).

Results

Comparison of differences between the 1975 and 1979 trainee soldier samples involves much more than the analysis of the time difference. First, the two samples represent potentially different loci of the several dimensions analyzed in this investigation. In addition, the data collectors were not the same for the two projects. The two samples involve substantially different kinds of MOS training which may conceivably involve different kinds of stress for the individual. Finally, the fact that the two samples involve vastly different MOS's (i.e., combat arms vs. medical) also introduces the possibility of differences due to self-selection factors, i.e., the type of person who enlists in the U. S. Army to become an infantryman may be substantially different from his medical counterpart.

These qualifications notwithstanding, comparison of the 1975 and 1979 samples is both useful and instructive. The 15 variables which constitute a common basis of comparison between the 1975 and 1979 samples are depicted in tabular form which reduces the 15 variables to three comparison profiles: (1) demographic; (2) family history; and (3) psychiatric. Perusal of the information in each of the three tables reveals that the results are mixed, i.e., while some variables are remarkably stable across time and between groups, others vary considerably.

The cluster of demographic and family history profiles depicted in Tables I and II indicate that the male soldier identified by the mental health facility as "Unsuccessful" in 1975 is remarkably similar in personality characteristics and background to his 1979 counterpart. As depicted in Table I, the 1975 and 1979 "Unsuccessful" soldier samples differ only .7 of a year in age; .1 of a year in the age at which schooling terminated; .4 of a year in educational level; and retains essentially the same distribution with regard to marital status. Table II indicates that the 1975 and 1979 "Unsuccessful" soldiers came from families of essentially the same size with approximately the same distribution with regard to family position. The unsuccessful soldier of 1979 reported a 5% decrease in the incidence of previous psychiatric care for a family member while, at the same time, exhibiting an 8% increase in the reported incidence of parental separation or divorce. Table III, psychiatric profile, is the area in which differences between the two samples are most clearly evident. The unsuccessful soldier of 1979 reported a 22% increase in the incidence of previous psychiatric care for himself. While reporting a 5% decrease in problems associated with maladjustment (e.g., anxiety, stress, and depression), the unsuccessful soldier of 1979 reported a 33% increase in overt discharge-seeking and a 6% increase in inability to relate to others, an incidence three times that of the 1975 sample. Finally, while the clinical impressions of the 1979 sample reveals a modest increase of 4% in the category of adjustment reactions, an increase of 13% was noted for the category of personality disorders.

Discussion

As pointed out by a number of previous studies, demographic profiles tend to be syndromatic, i.e., certain variables tend to be associated with

one another in clusters. Such clustering which has been evident for "Unsuccessful" soldiers in a number of previous studies, has probably also been evident for the successful trainee. Moreover, the successful trainee may share some important commonalities with his unsuccessful counterpart. Therefore, causal inferences cannot and should not be made simply by inspection of variable clusters. However, it can be stated with some degree of confidence that the results of this study constitute confirmation of the well-established proposition that individuals with poor pre-service adjustment patterns also tend to experience adjustment difficulty in the military. The evidence which emerged from this study indicates that while there is apparently a remarkable stability in the demographic and family history profiles of unsuccessful male soldiers between 1975 and 1979 a number of possibly significant differences were noted in the psychiatric profile. While it is possible that the two groups are not comparable samples, but reflect two entirely different populations, the commonality of demographic and family history profiles tends to lend less credence to this hypothesis.

What is discernible is a pattern of increased family dysfunctioning, fewer attempts to ameliorate disruptive events, and a higher incidence of individual psychiatric problems. Inside the military system, the unsuccessful soldier of 1979 reports less stress, but exhibits a much higher incidence of both an inability to relate to people in the system and the seeking of relief via discharge rather than attempt to adapt. Consequently, he is seen by military mental health counselors as presenting with a higher incidence of personality disorders (i.e., lifelong patterns of maladaptive behavior) and alienation which apparently preclude his successful adjustment to the military.

While inconclusive, these data are supportive of previous observations by Wesbrook and Reed which postulate patterns of dissatisfaction, social and political alienation which result in decreased combat effectiveness.

References

Carney, L. Trainee Discharge Rate Jumps. Army Times. June 18, 1979.

Reed, J. L. An Analysis and Evaluation of the United States Army (The Beard Study), in Status of the All-Volunteer Force. U. S. Senate Committee on Armed Services. U. S. Gov't Printing Office, June 20, 1978.

Wesbrook, S. D. The Alienated Soldier: Legacy of Our Society. Army. 1979, 29, 18-23.

Worthington, E. R. The American Soldier: Those who make it and those who do not. Paper presented at the Regional Conference of Inter-University Seminar of Armed Forces and Society, Arizona State University, Feb 1976

I Demographic Profile

Variable	1975 Soldier N=75	1979 Soldier N=16
Age	19.9	20.6
Educational Level	10.8	10.4
Age School Terminated	17.5	17.4
% with High School Diploma	49%	38%
Marital Status		
Single	71%	69%
Married	19%	19%
Div/Sep	9%	12%

II Family History Profile

Number of Siblings (Incl S)	4.7	4.5
Family position		
Oldest	23%	31%
Middle	41%	44%
Youngest	33%	25%
Only	3%	9%
Previous Family Psychiatric Care	43%	38%
Parents Div/Sep	42%	50%
Age at Div/Sep	8.3	10.6

III Psychiatric Profile

Week of Training presented at CMHS	2.9	4.6
Previous Psychiatric Care	15%	37%
Presenting Problem as seen by soldier		
wants discharge	19%	52%
Adjustment Reaction	44%	39%
Inability to relate to others	3%	9%
Somatic complaints	6%	0%
Clinical Impression by CMHS		
Adjustment Reaction	60%	64%
Personality Disorder	23%	36%
Other	17%	0%

Female Adjustment in The Volunteer Army

E. R. Worthington
Brooke Army Medical Center, Fort Sam Houston, Texas

Abstract

This study examined adjustment problems encountered by female soldiers (trainees and career military) assigned to a major Army post. Comparisons with male peers were also evaluated to ascertain if females differ in their methods of coping from their male counterparts. While differences were found between male and female trainees, they probably are more indicative of changing rules, regulations and sex roles than male-female coping patterns. Major differences were noted between career males and females; more females are single, more were self-referrals to mental health (males were more often referred by their unit or the hospital) and most females experienced adjustment reactions while males encountered more diversification in psychiatric problems. While their methods of coping might vary, males and females who can't adjust to the military are not that different.

The female soldier is viewed as a potential solution to the problem of the Army's inability to attract quality male recruits (Worthington, 1979). While all the services are increasing female recruitment; the question still remains as to what limits should be placed on total integration of women in the Armed Forces. Civilian and military leaders are not in agreement over the females' physical, emotional, and psychological ability to adapt and perform in combat operations. Research on this subject is limited to rare factual instances in which females have engaged as combatants or anecdotal reports.

Presently it is not possible to seek this information except under simulated conditions (Marlowe, 1978) or by comparing females with males, both whom are sharing the same experiences. Comparisons may be made on how each sex adjusts to the stresses prevelent within the same environment. These data may be helpful in predicting the female's performance under combat conditions.

This study examined adjustment problems encountered by female soldiers (trainees and permanent party) assigned to a major Army post. Comparisons with male peers were made to determine if females differ in their method of coping.

Method

During the summer of 1979, there were 224 initial male and female soldier visits to the Community Mental Health Service. These 224 soldiers included 148 trainees and 76 permanent party. Among the trainee personnel 106 were male and 42 were female; the permanent party soldiers were comprised of 62 males and 14 females. Of these 224 soldiers, 16 males and 7 females were recommended by CMHS for immediate discharge due to their inability to adapt to the military environment. Demographic, family and psychiatric profiles are depicted in Tables I, II and III.

Results

Trainees: The female is younger and better educated than the male. More females are single and divorced than males. While the majority of the male trainees are middle children the female trainee is almost equally

distributed within the sibling birth positions. The educational advantage the female enjoys is most likely related to regulations which required the female enlistee to possess a high school diploma. Significant differences are noted in the psychiatric profiles. Females have a lower incidence of pre-service psychiatric care and remain in training 1½ weeks longer than males before presenting to CMHS. The women request discharge at a rate 39% higher than males yet their claim of not being able to relate to others is 31% less. More adjustment reactions and fewer personality disorders are noted in female trainees. They seek psychiatric help more than males who are most often referred by their unit or from another part of the hospital. More females return for treatment while more males are only seen once (reflecting the higher referral rate, usually for evaluation for administrative disposition rather than treatment). The percentage of trainees recommended by CMHS for discharge is essentially the same for both sexes.

Permanent Party: While the age separation (females younger) becomes more apparent, the education level is similar with 93% of the females and 90% of the males being high school graduates. Marital differences are significant. The females are fairly evenly distributed between single, married and divorced or separated categories, but single males represent only 18%, with 61% being married and 21% being divorced. More than half (56%) of the females have a family history of psychiatric care yet only 28% of the females come from a broken home. The males report the reverse with 32% having a family history of psychiatric care and 50% coming from broken homes. The males presented to CMHS with a wide variety of complaints but only 5% specifically sought separation. Almost 40% of the females complained of adjustment problems and 28% wanted a discharge. Clinically, 91% of the women were seen as encountering adjustment reaction problems while this accounted for only 50% of the males. Males were evaluated as having a much broader spectrum of clinical diagnoses. As in the case of female trainees, most of the women were self-referred (64%). Only 41% of the males were self-referred with 29% unit referrals and 21% hospital referrals. More than half of the both sexes (females 53% and males 62%) returned for follow-up treatment. More females (13%) were seen as unable to adjust and recommended for discharge than males (2%).

Recommended for discharge: The most prevalent female complaint (63%) was dissatisfaction and a desire for discharge (males were 52%). The complaint of maladjustment was three times as high for males (39% vs 13%). The clinical diagnoses for both sexes was similar, two thirds were diagnosed as adjustment reaction and one third personality disorder. Most of the soldiers were unit referrals (females 80%, males 76%); 18% of the males were self-referred but no females were, the other 20% of the females were medically referred. While 40% of the females returned for help before discharge, only 14% of the males returned.

Demographically the differences between these males and females are similar to the differences noted among the trainees. This female though, is slightly older than the male but less educated than other female soldiers (80% high school diploma), more females are single (80%) and more are the youngest child (60%).

Discussion

Due to societal shifts in sex-roles (ISR Newsletter, 1980) and the changes in recruiting policies for females (Finegan, 1979) the comparisons

between male and female trainees probably does not reflect much more than temporary differences.

There are some notable differences which suggests that the sexes do vary in their methods of coping. These data also reveal other male-female differences. As the soldiers progress in the military, males become married and remain in the Army; females, tend to leave the service after marrying. While there is no predominant psychiatric diagnostic category for males, females are most often seen as experiencing an adjustment reaction. Possibly this is associated with their motivation to seek help while males are reluctant to recognize emotional conflicts. While the percentage of females seeking help is higher, so to is their perception of the solution being a discharge.

While their methods of coping might be different, males and females who cannot adjust to the military are not seen as that different.

References

Finegan, J. Army Reports Sharp Dip in Women Recruit Quality. Army Times. 24 December 1979.

ISR Newsletter. Fifteen-Year Study Documents Tremendous Change in Women's Sex-Role Attitudes. University of Michigan. Winter 1980.

Marlowe, D. H. Military Preventive Psychiatry. Walter Reed Army Institute of Research, Washington, D. C. 1 October 1978.

Worthington, E. R. The Economics of Human Resources in the U. S. Army. Paper presented at Inter-University Seminar of Armed Forces and Society. USN Postgraduate School, Monterey, California, May 1979.

Variable	Female Trnees N=42	Male Trnees N=106	Female PP N=14	Male PP N=62	Female* RFD N=5	Male RFD N=16
I Demographic Profile						
Age	19.6	21.5	25.2	29.7	22.2	20.6
Education Level	12.4	11.3	13.0	12.9	13.0	10.4
Age School Term	18.5	16.8	18.5	18.6	17.8	17.4
% with H.S. Diploma	93%	58%	93%	90%	80%	38%
Marital Status						
Single	70	68	36	18	80	69
Married	17	22	29	61	20	19
Div/Sep	13	9	36	21	0	12
Widow	0	1	0	0	0	0
II Family History						
Number of Siblings	5.1	5.9	5.0	5.1	4.4	4.5
Family position						
Oldest	35	24	38	35	20	31
Middle	37	56	46	44	20	44
Youngest	26	17	15	19	60	25
Only	2	3	0	2	0	0
Prev Fam Psych Care	37	35	56	32	20	38
Parents Div/Sep	52	54	28	50	40	50
III Psychiatric Profile						
Prev Psych Care %	29	34	21	27	20	37
Week of Training Complaint Presented	5.4	4.1	N/A	N/A	N/A	N/A
Psychiatric Complaint						
Desires Discharge	43	31	28	5	63	52
Adjustment Reaction	31	26	39	28	13	39
Inability to Relate to others	9	13	11	27	13	9
Somatic Complaints	2	4	0	5	0	0
Other	15	26	22	35	13	0
Clinical Diagnosis						
Adjustment Reaction	71	62	91	50	64	66
Personality Disorder	20	29	9	17	36	33
Other	8	9	0	33	0	0
Referral Source %						
Self	50	42	64	41	0	18
Unit	26	33	21	29	80	76
Medical	20	25	14	21	20	6
Other	4	1	0	9	0	0
CMHS Disposition %						
RTD w/o RTC	37	41	33	28	60	86
RTD w/RTC	47	43	53	62	40	14
Refer out	7	6	0	8	0	0
Recommend for Discharge	5	15	2	1	0	0

*N=5 (2 RFD were deleted from study due to lack of data)

Some Army Psychologists Remain In The Service - Why?

A. David Mangelsdorff, PhD
Health Care Studies Division
Academy of Health Sciences (HSA-CHC)
Fort Sam Houston, Texas 78234

Responses of Active Duty Army psychologists who had participated in a survey in December 1976 were compared with the responses of psychologists who had left the Army. Regression models were developed to predict "Likelihood Extend" and "Likelihood Remain Until Eligible to Retire."

The retention of health care professionals like psychologists and physicians has been a continuing concern for the military services. The problem of retaining military medical officers has been addressed at varying times (Baker, 1969; Boyson, 1967; Braunstein, 1974; Cooke and Mixson, 1967, 1971; Dully, 1974; Hedlund, 1968; Jorlette, 1975; Krause, 1978; Lanier, 1975; Mangelsdorff, 1978; Mangelsdorff and Hubbart, 1976; Murray, 1978; Winkler, 1968). Among the recurrent findings influencing health care professionals to leave the service were: inadequate pay, possibility of command or administrative assignments, lack of amount of participation in making decisions affecting own career, lack of sense of belonging to the community and social life of the military, and poor facilities.

The intent of this study was to examine retrospectively the responses of Army psychologists who had responded to a survey conducted in late 1976-1977 (Mangelsdorff, 1978). The purpose was to document what factors induce some psychologists to remain in the Army and what cause other psychologists to leave the service. Specifically, what factors increase the likelihood of psychologists (1) extending beyond their current obligation and (2) remaining on active duty until eligible for retirement.

Method

Subjects. Psychologists who were on Active Duty in the Army as of November 1976 (N = 130) and psychologists who had left the Army since July 1974 to November 1976 (N = 69) were included.

Procedure. In December, 1976, each psychologist was mailed a survey instrument which requested demographic background, military background, attitudes toward military career, levels of satisfaction, retention factors, and the Job Descriptive Index (Smith et al, 1969). Psychologists who had left the Army between July, 1974 and November, 1976 were instructed to answer the survey questions as though they had six months remaining in their active duty obligation.

Active Duty Sample Regressions. Stepwise regression analyses were performed to predict responses to each of the 7-point criterion items: (1) "Likelihood extend beyond current obligation" and (2) "Likelihood remain until eligible for retirement" (1 = low probability, 7 = high probability) using as independent variables the responses by the Active Duty sample (N = 76) to Long Term Motivator set attitude items (7-point Likert), Demographic section items, and the Job Descriptive Index scales. The raw beta weights developed from the Active Duty (AD) sample were applied to the responses of the Left Service (LS) group and to the Active Duty Left Service (ADLS) group.

Results

Sample Characteristics. Of the Active Duty psychologists, 114 of 130 (88%) responded; of the psychologists who left the Army, 48 of 69 (70%) returned the questionnaire. The Active Duty sample was further divided into Army psychologists who remained on active duty (AD) through January 1980 (N = 76) and those who left the Army (ADLS) during the time period November 1976 to January 1980 (N = 38). Of the 16 Active Duty psychologists who did not respond, eight left the Army. The average age of the sample was 31 years.

There were no significant differences between groups as a function of age. There was a significant difference for Total years of active military service completed ($F = 3.60$ ($df = 2/159$, $p = .029$), with the AD group having significantly more years completed than the LS group. For Years of prior active military experience before becoming a psychologist, there was a significant difference between groups ($F = 6.70$ ($df = 2/158$), $p = .001$), with the LS group having significantly fewer years than either the AD or ADLS groups.

Job Descriptive Index. There was only one significant difference between Active Duty versus Left Service and Active Duty Left Service groups on the JDI scores, that for PROMOTIONS, where the AD group reported significantly greater satisfaction than either of the LS or ADLS groups ($F = 8.97$ ($df = 2/157$), $p = .0002$).

Validation Samples. Regression equations to predict responses to each of the 7-point criterion items: (1) "Likelihood extend" and (2) "Likelihood remain until eligible for retirement," using the raw beta weights developed from the Active Duty sample were applied to the responses of the Left Service group (N = 48) and to the Active Duty Left Service group (N = 38). Tables 1 and 2 summarize the comparisons.

Remain in Service. A correlation matrix was developed to display the relationships between the JDI subscales, Age, Rank, Total years active military service completed, Years prior active military experience before commissioning as a psychologist, Sense of membership in Army, Personal accomplishments as a military psychologist, Likelihood promoted, Likelihood remain until eligible to retire from military service, the regression equation predictions for retire and extend, and a factor Remain in Service. The factor Remain in Service assigned a value of one to all Active Duty psychologists and a zero to all ADLS and LS group members. Table 3 depicts the matrix.

Discussion

The correlations between each of the criterion variables and the factor Remain in Service were both highly significant (Likelihood Extend: $r = .61$; Likelihood Retire: $r = .60$; $p < .001$). These findings support the components of the withdrawal decision process offered by Porter and Steers (1973), Locke (1976), and Mobley (1977, 1979). Mobley et al (1978) reported for precursors of hospital employee turnover that intention to quit was a significant predictor of actual attrition.

In the present study, Sense of membership in Army was one of the variables with the highest correlation with the factor Remain in Service ($r = .40$, $p < .001$). In comparing the Active Duty versus both the Left Service groups for responses to Sense of membership in Army, those leaving the

service reported a minimum. The Total years of active military service completed (the highest beta in the analysis) was the best predictor of the criterion "Likelihood remain until eligible to retire." Total years of active military service was moderately related to the factor Remain in Service ($r = .19$, $p = .007$). Since most psychologists enter the Army as Captains and many leave the service at the same rank as they entered, there may have been some feelings of powerlessness in their position felt by some of the psychologists who left the service. Shephard and Panko (1974) note that power-deficient workers had less commitment to organizational goals. Of the Left Service and the Active Duty Left Service groups, almost all rank ordered themselves as psychologists first, and most placed their rank ordering of self as military officer as lowest.

Issues which contribute to categorizing a psychologist in the Active Duty group dealt with the development of an identity as a career military officer. Specifically, the satisfaction with PROMOTIONS significantly separated those psychologists who left the Army from those who remained on Active Duty ($p = .0002$). The two items having the highest beta weights in the regression for Likelihood Extend were: Sense of Membership in Army and Personal accomplishments as a military psychologist. This supports the notion of developing the identity of a career military psychologist. Personal accomplishments as a military psychologist was significantly related to the factor Remain in Service ($r = .24$, $p = .001$).

With competition from civilian jobs perhaps offering higher pay, independence, stability, opportunity for self-improvement, or some facet not otherwise found in the military, an Army psychologist may choose not to remain in the Army. In addition, factors supporting a negative view of the military may influence the decision to leave the service.

The benefits of joining the Army for educational opportunities must also be emphasized. A sizable number of psychologists join the Army for the Graduate Student Program and for the educational benefits available after service. In predicting the criterion Likelihood Extend, a large negative weight was assigned to Having opportunity to receive post-doctoral training, implying there were few opportunities available in the military. If more opportunities for professional growth and development were available through the Army, perhaps more psychologists would consider remaining on active duty to take advantage of them. If the organization can provide the opportunities for its personnel to meet their needs, the likelihood of turnover would be reduced. For Army physicians, the availability of residencies and post-residency specialized fellowships has been cited as a very significant factor in why some physicians remain in the Army (Krause, 1978; Whelan, 1974).

Footnote

[1]Health Care Studies Division, Academy of Health Sciences, Fort Sam Houston, Texas, 78234, is the work address.

References

Available on request.

Table 1

Validation Sample Using Regression Weights Developed From Active Duty Sample To Predict "Likelihood Extend Beyond Current Obligation" For The Left Service And For The Active Duty Left Service Groups

Active Duty Left Service Group

		Actual probability extend		
		Low	Undecided	High
Predicted	Low	7	2	0
Probability	Undecided	9	15	2
	High	1	2	0

χ^2 = 5.66 (df = 4), p = .225
Pearson's r = .302, p = .03

Left Service Group

		Actual probability extend		
		Low	Undecided	High
Predicted	Low	11	0	0
Probability	Undecided	26	4	2
	High	4	0	1

χ^2 = 4.55 (df = 4), p = .335
Pearson's r = .223, p = .063

Note: Both the "Actual" and the "Predicted" score using the Active Duty regression weights were collapsed as follows: 1, 2 = Low; 3, 4, 5 = Undecided; 6, 7 = High.

Table 2

Validation Sample Using Regression Weights Developed From Active Duty Sample To Predict "Likelihood Remain Until Eligible To Retire" For The Left Service And For The Active Duty Left Service Groups

Active Duty Left Service Group

		Actual probability retire		
		Low	Undecided	High
Predicted	Low	14	2	0
Probability	Undecided	10	10	1
	High	0	0	1

χ^2 = 24.71 (df = 4), p = .0001
Pearson's r = .521, p = .0004

Left Service Group

		Actual probability retire		
		Low	Undecided	High
Predicted	Low	23	3	0
Probability	Undecided	16	5	0
	High	0	0	1

χ^2 = 49.26 (df = 4), p = .0001
Pearson's r = .357, p = .006

Note: Both the "Actual" and the "Predicted" score using the Active Duty regression weights were collapsed as follows: 1, 2 = Low; 3, 4, 5 = Undecided; 6, 7 = High.

Table 3

Correlation Matrix of Factor Remain in Service With JDI Subscales,

Demographic and Attitudinal Responses

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1																
2	04															
3	01	16														
4	27	30	22													
5	15	51	14	14												
6	07	43	11	17	32											
7	-00	10	05	06	04	02										
8	16	05	01	19	03	-00	77									
9	15	05	-04	09	06	-00	86	76								
10	23	00	04	19	13	-01	48	43	56							
11	38	13	01	33	18	13	43	45	47	43						
12	15	01	01	25	03	09	22	18	23	26	41					
13	29	07	-05	43	15	21	-23	-13	-18	03	17	05				
14	59	18	12	40	41	14	33	41	43	30	60	45	29			
15	57	15	05	40	22	10	42	47	57	38	68	45	29	87		
16	26	15	06	36	19	12	60	54	66	45	69	66	20	68	75	
17	24	15	08	38	17	11	56	50	60	38	68	68	20	67	73	96

Variables:

1	Remain on AD	9	Total years of active military service conpleted	16	Extend
2	JDI work	10	Years prior active military experience	17	Retire
3	JDI pay	11	Sense of membership in Army		
4	JDI promotions	12	My personal accomplishments as a military psychologist		
5	JDI co-workers	13	Likelihood promoted		
6	JDI supervisor	14	Likelihood extend beyond obligation		
7	Age	15	Likelihood retire from military service		
8	Rank				

Notes: 1. Decimal points omnited in matrix.
2. The factor Remain in Service assigned a value of one to all AD psychologists and a zero to al AD-LS and LS group members.

Clinical Psychology in the Air Force: A Continuing Struggle for Parity

Donald E. Trahan
Keesler USAF Medical Center
Dale P. Mann
Peterson USAF Clinic

Abstract

This paper reviews a number of perplexing problems which psychologists have faced in their attempts to gain parity with medical professionals in the field of mental health care. Differences in the allotment of constructive service credit for pay and rank, professional pay, and numerous other benefits have created a situation in which clinical psychologists are relegated to a role as "second-class citizens" in the field of mental health care. It was suggested that elimination of these inequities could not only improve the morale and satisfaction of psychologists as a group, but also could significantly enhance the ability of the Air Force to retain professionals in this crucial field as well as ultimately improving the overall quality of mental health care provided within this branch of service.

During the past decade the status of Clinical Psychology as a profession has seen some significant changes. During these years psychologists have been recognized as independent providers of mental health care. They have come to play an important role in virtually every phase of mental health programming in this country. As others have become aware of the high level of training and expertise of the clinical psychologist, the profession has gained increased legal recognition. In fact, psychologists now are recognized as independent providers of health services by 80% of the private health insurance companies in this country, as well as by many governmental health agencies.

Despite these positive changes, the struggle of Clinical Psycholigists for parity with other mental health professionals has not been an easy one. There have been continued efforts by various groups of medical professionals, particularly psychiatrists, to deny psychologists the right to provide essential health care services, and generally to relegate psychologists to a role as "second-class citizens" in the field of mental health care. These efforts have included not only preplanned, well-organized, and well-financed efforts in the courts, but in other instances also have included personal attacks upon psychologists as individuals and as a group.

These confrontations between Psychiatry and Clinical Psychology have been prominent not only in civilian settings, but they also have caused considerable concerns in the military services. These concerns have been voiced repeatedly by Air Force Psychologists who have experienced significant problems in these years due to what they perceive as the preferential treatment being given to psychiatrists and other medical professionals. Since the problems facing Air Force psychologists are essentially the same as those experienced in other branches of the military service, this review will focus exclusively upon issues presented by Air Force psychologists.

Within the Air Force the attrition rate for Clinical Psychologists has increased considerably when compared with other specialties in the

BEST COPY AVAILABLE

Biomedical Science Corp. In fact, the turnover rate now approaches 100% for any given 4-year period. Clinical Psychologists repeatedly have stated that the reasons for this attrition rate are quite clear to them. Psychologists as fully qualified practitioners in the field of mental health care provide a variety of services. These include, among many others, applying psychological theories, principles, and techniques in providing direct patient care; consultation on a continuing basis with other medical professionals; participation in the education and training of hospital personnel, including many medical residents; participation in research on problems in human effectiveness and emotional disturbance in medical and other settings; assessment and disposition of cases involving character and behavior disorders; and participation in preventative health care programs. Not only do psychologists provide essentially the same services as their psychiatric colleagues, but in many cases they provide services which psychiatrists are unable to provide either because of their lack of training or their legal status. The most obvious of these services include psychological assessment, including both objective and objective testing; standardized intellectual assessments; and neuropsychological services. In addition, by virtue of their training, psychologists are providing the majority of services in a number of other essential areas. In fact, recently several psychologists have been noted for their outstanding work in the areas of stress-management, biofeedback, research, and preventative health care.

Despite these very significant contributions, psychologists within the Air Force have continued to be placed in a role as second-class citizens in the field of mental health care. Although the clinical psychologist fulfills a role at least equivalent to that of any other mental health provider, he has not been afforded the same benefits or even considerations as these other providers, particularly psychiatrists. The reasons for the continued disparities are rather complex, but will be covered in detail in the remainder of this paper.

Issues of Disparity

Constructive Service Credit:

Constructive service credit is defined as the number of years credit in the military service for each year of education, internship, residency, experience, and several other factors. The reason that constructive service credit is such an important issue is that, at least for psychiatrists and other physicians, it determines both rank and pay at the time one enters the service.

A fully qualified psychologist enters the service at the rank of Captain by virtue of the amount of credit he has been given for his education. However, even if additional education, training, and experience have been obtained, no further credit is given. On the other hand, psychiatrists receive credit for their medical school training, internship, residency, experience, medical board certification, and even education in a second health discipline, above and beyond the rank of Captain, all the way to the rank of full Colonel. In fact, it is extremely common for physicians to enter the service at the rank of Major or Lt. Colonel, and some have entered as full Colonels. As one can clearly see, there are very significant disparities in the amount of credit allowed to psychologists and psychiatrists. As one also can see, psychologists are at a very clear disadvantage from the very beginning in terms of rank within departments of

mental health.

Another crucial element in the constructive service issue is the effect such allowances have on pay. Since salary is determined by both rank and years in service, allowing credit for time in service leads to pay inequities above and beyond those resulting from differences in rank. In the Air Force, psychologists receive absolutely no constructive service credit for pay purposes. On the other hand, psychiatrists, even those just out of school, receive a minimum of 4-6 years credit for pay purposes alone. When combined with the credits given for rank, this creates a very substantial difference in pay between the psychologist and psychiatrist. However, there are other factors which increase these inequities even further, and these will be examined in the next section.

Professional Pay

Professional pay is yet another factor which has caused considerable dissatisfaction for clinical psychologists. Like other benefits, psychologists receive absolutely no professional pay. In contrast, psychiatrists and other physicians may receive up to $350.00 per month in professional pay above and beyond their normal salary and benefits. Needless to say this enhances pay inequities even more. It is apparent that the "equal pay for equal work" philosophy is not being adhered to in the field of mental health care. The Air Force has contended that such pay is necessary to attract and retain professionals who otherwise would be in short supply. The implication is that these recruiting problems do not exist for Clinical Psychology. However, data consistently support the fact that there is very little difference between the manning postures for Psychiatry and Clinical Psychology. Figures for both 1978 and 1979 reveal that the number of slots not filled for the two professions differed by only a few. Even more revealing was the statement by Lt. Colonel H. Wagner, Consultant to the Surgeon General for Psychiatry. He indicated that figures for Psychiatry indicated "no personnel problems", with approximately 90% of their positions filled.

When all this had been revealed in several forums, including Congressional subcommittees, The Air Force apparently changes its position somewhat, saying it is the shortage of physicians in general which justified professional pay. They stated further that in time of war psychiatrists could serve as general medical officers - something the Clinical Psychologist could not do. Psychologists have responded by pointing out that our country has been at war for only a small portion of the past century and that the need for ongoing professional pay for psychiatrists on this basis, while excluding psychologists, simply is not justified. In addition, this line of reasoning totally overlooks the fact that psychologists routinely are trained in many areas in which the psychiatrist has little or no training. Typically, psychiatric training is limited primarily to medical and traditional analytic approaches. Yet, the psychologist is not paid a single cent for the services he provides above and beyond those of the psychiatrist.

Promotion Policies

Another area of obvious inequity for the clinical psychologist has to do with promotion policies. To be frank, promotional opportunities for psychologists are absolutely dismal when compared with those of psychiatrists, and even when compared with many specialties within the Biomedical

Science Corp. Under Air Force guidelines 100% of physicians (including Psychiatrists) are promoted to the rank of Major within six years, 90% to the rank of Lt. Colonel within 12 years and 80% to the rank of full Colonel within 18 years. These figures do not even reflect the fact that a significant portion of physicians may enter the service at a rank higher than Captain. This also does not account for the fact that many medical officers may be given "early" promotions, without having their promotions counted against the primary zone percentages.

Clinical Psychologists in contrast have a separately derived promotion schedule which is considerable more stringent than that of Medical Corp. officers. For the psychologist, promotion to Major typically takes 9-11 years; and then only 80% can be promoted. It generally takes 15-17 years to make Lt. Colonel, and at best only 70% can be promoted. In theory, a psychologist has a 50% chance of making full Colonel in 21-23 years; in reality, very few actually are promoted to that level. As one can clearly see, even the regulations are highly inequitable. However, the clearest indications of the present disparity is to look at actual data. The most recent figures available to this author are presented below. They are current as of May 1979 and are entirely representative of more recent data.

	Clinical Psychology	Psychiatry
Colonel	1	14
Lt. Colonel	8	46
Major	13	41

It is important to realize first that the number of Clinical Psychologists and Psychiatrists in the Air Force is approximately the same. Second, in view of this data, the theoretical promotion rates for psychologists are irrelevant. Over the years, promotion rates for psychologists have fallen well below the theoretical rates.

After these matters were investigated by a Congressional subcommittee last year, the Air Force responded by denying the inequities overall, and stating that during the last few promotion boards psychologists have not fared badly, at least when compared with other Biomedical Science officers. Although this may be true, these boards had not taken place until after considerable publicity had been given to the matter at a Congressional level. However, those recent figures do not reflect the actual promotion rates of the past 10-15 years, which reveal obvious inequities. In addition, figures from the last promotion boards indicate the promotion rates for psychologists have once again returned to their previous levels.

Department Chairmanships

It is the Air Force policy that clinical psychologists are not allowed to function as chiefs of mental health departments except in those very rare situations where either a psychiatrist or a social worker is not assigned to the base. In reality, this relegates the clinical psychologist to having the dubious distinction of being chiefs of one-man units. Even in these few cases, the psycholigist is usually responsible to a physician. This is truly remarkable in that most physicians have had at best a few introductory courses in psychology. In addition, this situation is at times inconsistent with the training, experience, and years of service of the individuals involved.

The Air Force policy on this matter, which is heavily reinforced by psychiatrists, is at best outdated. Their reasoning is that psychologists should be supervised in all cases by physicians, since part of the work involves differentiating physical from emotional disorders, and that psychologists are not qualified to assume responsibility for medical patients. However, there are several important aspects of this issue. First, whenever this type of diagnostic dilemma arises, the patient typically is referred to a clinical psychologist for differential diagnosis. In reality, it is the clinical psychologist who deals with this problem on a day-to-day basis. Furthermore, there is no reason to believe that the physician of psychiatrist is more competent to deal with this problem than the clinical psychologist. Although Medical Corp. officers unquestionably have superior training in medicine, the clinical psychologist in the most highly trained mental health professional.

Perhaps even more important is the fact that these issues of patient care have nothing whatsoever to do with qualifications for departmental chairmanships. However, psychiatrists typically rely upon these sorts of arguments as justification for maintaining their privileged positions.

It seems surprising that the Air Force has continued to support such an outdated position despite national and worldwide recognition of clinical psychologists as fully autonomous mental health providers. All 50 states have passed legislation allowing clinical psychologists to function as health care practitioners independent of the supervision of either a psychiatrist of other physician. Not only has this separation been recognized by most insurance companies. but various governmental agencies such as the Veteran's Administration have even gone so far as to establish separate departments of psychology within their hospitals. In addition, many state and local hospitals, as well as many numerous medical training facilities, presently employ clinical psychologists not only as providers of health care services, but also as teachers in the field of mental health care. From these examples, it is very clear that clinical psychologists have established themselves a professionals apart from Psychiatry. Psychologists have their own professional organizations, standards, and ethics. In fact, their range of training and experience far exceeds that of any other mental health professional.

Despite all this, the Air Force has continued to support its outdated philosophy, to the merriment and satisfaction on most psychiatrists. This does seem rather ironic in view of the fact that individuals as high as the Deputy Assistant Secretary of Defense have suggested that psychologists be allowed to function as the head of a mental health department.

Conclusions

As is evident from this review, clinical psychologists in the Air Force face a number of very perplexing problems. These problems not only interfere with the most appropriate use of their skills, but they have led to serious dissatisfaction within their ranks. Although these problems could be resolved through military channels, the Air Force has shown little interest in initiating any positive changes; in fact, no action whatsoever has been taken. With input coming almost exclusively from psychiatrists and other physicians, The Air Force has staunchly defended these inequities despite considerable pressures from members of both the Senate and House of

Representatives during the past year. Quite frankly, it appears that little will be done to improve the plight of psychologists as long as the Air Force relies upon physicians for its primary source of input. However, considerable interest has been shown for these issues by a number of key members of the House and Senate. These particular Congressmen seem convinced that positive changes can be made which will not only eliminate the numerous inequities inherent in the status quo, but ultimately also will improve the overall quality of mental health care provided by the military services. There are several key changes being recommended. First, psychologists should receive constructive service credits for promotion and pay for their education, training, and experience in a manner similar to that now credited to psychiatrists. Second, professional pay should be extended to psychologists as it is to psychiatrists. Third, psychologists should be allowed to function as chiefs of any Department of Mental Health whenever they are senior officers, regardless of other manning within the department. Ultimately, these changes could produce a number of distinct advantages for the Air Force as well as clinical psychologists. Eliminating inequities in pay, rank, and numerous other benefits would substantially improve morale and satisfaction of an entire group of professionals within the Air Force. These changes could substantially improve the Air Force's ability to retain qualified professionals within the field of mental health care. This would represent a significant change form the status quo where over 81% of the active-duty psychologists are Captains. However, it is evident that none of these changes will come easily. Any change at all will require a unified effort by all military and civilian psychologists. In addition, it is apparent that Congress must exert considerable pressure before the issues will even be seriously considered. However, with a sustained effort by all these groups, perhaps the future will bring an end to these long-standing and system-wide disparities.

A New Approach to Undergraduate Leadership Education at the Air Force Academy: Combining Psychology with History

Lawrence F. Silverman
and
Fred D. Harburg

Abstract

This paper describes a new approach to the study of leadership conducted at the United States Air Force Academy. This approach is based on our conclusion that managers and leaders are fundamentally different personalities that develop in quite different ways. We take the position that the proper study of leaders is leaders. We have combined the study of the acts of historical personages with the study of their psychological make-up.

We are made aware of the critical need for capable leaders by today's bewildering complexities and our inability to deal with them. Companies, countries, and organizations of all kinds seek to attract and develop people who can deal effectively with the critical problems of the present age. This paper describes a new approach - although it has elements that reach back to antiquity - to the study of leadership conducted at the United States Air Force Academy.

Like many organizations, military and otherwise, the United States Air Force has developed an educational and training program designed to prepare leaders. The first stage of this program occurs during the four year undergraduate schooling of officer candidates at the United States Air Force Academy.

At the Academy, academic leadership training begins during the sophomore year when the cadet receives exposure to psychological explanations of the leadership process. This exposure includes the discussion of topics such as motivation, human relations, power and attitudes. The cadet also studies several leadership models for contemporary managers including those developed by Vroom and Yetton, Fiedler, Mitchell, Hersey and Blanchard and other well-known management theorists and researchers.

These are management/business models not unlike those used by contemporary American business schools and designed to fill the needs of the business world. Their study constitutes a sensible and useful experience which gives cadets at the Academy guidelines and reinforcement in dealing with the human problems of organizational behavior. If lucky, the cadet will recall some of the theories and models he/she encountered when dealing with the problems of managing people in the Air Force after commissioning. Those majoring in management or organizational behavior receive additional iteration of these approaches throughout their cadet careers and all cadets receive reexposure during their management courses, military science courses and cadet training. It is our opinion that repeated indoctrination in the management/business models of leadership dominates contemporary thinking about leadership in military as well as civilian life, and detracts from the need to study other kinds of leadership necessary to the operation of military and civilian organizations.

It is clear that there is more to performing the leadership role than allocation of resources or, as the editors of Harvard Business Review put it, maintaining the balance of operations (Zaleznik, 1977:67). What is missing is the creative, risk-taking, intense, empathic and occasionally charismatic kinds of activity associated with the attraction and persuasion of followers. The business model appears to minimize the question of how to effect change of direction for an organization. If change must take place, it is frequently directed by consultants or outside change agents. We believe change in the face of crisis is the central task of leadership. For that reason, that is, because we are interested in providing knowledge and insight into leadership for change, we offered a different approach to the study of leadership.

In constructing the course we found it essential to deal with the differences between managers and leaders. Here we drew on the work of Abraham Zaleznik of the Harvard Business School, and, in particular, upon an article he wrote which appeared in the 1977 May-June issue of the Harvard Business Review (Zaleznik, 1977:67-78). Others have also touched upon such differences, notably Eugene Jennings in his small book, The Anatomy of Leadership (Jennings, 1972), and Wrapp (1979) in his comments contrasting general managers and professional managers. Our conclusion is that managers and leaders are fundamentally different personalities who look at the world in different ways. For example:

Managers tend to view goals impersonally because their goals are not of their own creation but are derived from the organization. Leaders, on the other hand, are intensely involved in goals because the goals are their own. Managers view work as a problem-solving process and mandane work as necessary for survival. Leaders view work as a means to the realization of a vision. Managers view those they work with in terms of how they relate to the decision-making process and view themselves as well-socialized personalities with a stake in maintaining the current state of affairs. Leaders think of others in terms of what they want from decisions and of themselves as unsocialized, apart, and certainly different. We agree with Zaleznik (1977:70-74) when he equates management with stability and the maintenance of the balance of operations as the manager's primary function. We also agree that leadership implies instability albeit purposeful, vision-oriented change.

We take the foregoing as an assumption, that is, that leaders and managers are fundamentally different personalities. We also believe that history abounds with examples of leaders, the elements of whose personalities are more or less known to us, who operated in a wide manner of styles. It is logical to combine the study of the acts of historical personages with the study of their psychological makeup. This is what we have done, without, we hasten to add, attempting to place them upon the couch.

We take the position that the proper study of leaders is leaders and that the proper study of leadership rests on an examination of their acts. We are also convinced that what is typical of great leaders is also characteristic of leaders of small causes and small groups. Vision, empathy, risk-taking, good communication and a flexible accommodation to a variety of audiences are just as vital to the leader of a family as they are to leaders of armies, large corporations, or great social movements. We

believe, moreover, that students in leadership courses need to learn not only about qualities of personality and patterns of behavior, but also about the varieties of situations in which leaders emerge. A final consideration, one we are not likely to assess for a while, although we have some tentative indications from written student reactions to the course, is to help students sensitize themselves to elements of their own personalities and begin to make some judgments as to whether they are suited for careers mainly as managers or leaders.

As a theoretical basis for studying the formative influences upon leaders, we used the writings of Erik Erikson (1962, 1970:33-68) on psychohistory, specifically his study entitled "In Search of Gandhi," which is contained in one of our texts: Philosophers and Kings: Studies in Leadership, edited by Dankwart Rustow.

We then immersed the students in the reading of some fifteen or so biographical essays. These were written by classical as well as contemporary authors. For example, from Plutarch's Lives the students read selections on Alexander the Great and Julius Caesar. From Einhard (1977), a ninth century monk, they read about his master, Charlemagne. And, from a host of Ericksonians, they read about deGaulle, Nkrumah, Jamal ad-Din, Ataturk, Lenin, Isaac Newton, William James, and others (Rustow, 1970). In addition, the students viewed films of Churchill and Kennedy.

We ranged broadly through political leadership, military leadership, intellectual leadership, the leadership of social and religious movements, reform leadership and revolutionary leadership. In addition, a number of sessions were devoted to the nature of charismatic leadership, heroes and heroism, and the issue of whether and how organizations can develop leaders.

The classroom time was largely spent in discussion although the examination of each discreet figure was preceded by lectures, placing the leader in an appropriate historical and biographical setting.

The reaction of students (who were asked to provide anonymous responses to the course upon its conclusion) was positive, despite heavy and difficult reading assignments, essay examinations, and a term paper. Let them speak for themselves:

> ". . . I now appreciate the individual who is different and will try to lend that person my support and my encouragement especially when he may feel he has no one behind him."
>
> "The class makes me more perceptive (in) noticing the qualities of leadership in myself and others."
>
> ". . . I have seen in myself many of the qualities we talked of and have noticed those that I lack. . . Also, I have learned that it is important to support our leaders when they make mistakes and when they do well. . . Having had the course I will be less resistant to try my hand in leading. . ."
>
> "I think it (the course) made me aware of many of my own traits and qualities. . . This course has made me sit down and think about myself as a leader. I ask myself about my qualities (passionate? iron-willed?

pigheaded describes me better -- empathic? magnanimous?)
I also search for conflicts in my life."
". . . I feel that I am more aware of what a leader truly is and what traits a leader exhibits...maybe I will be more trusting of leaders and more willing to be led."
"I have also realized that leaders must be encouraged, and therefore I will do my best to avoid stifling other potential leaders..."
"I've been searching in myself to see if any of these qualities exist inside."

We would be less than candid if we did not report that some students, especially at the outset, strongly disagreed with our assumptions because they found it difficult to associate psychological conclusions or speculations with their own view of life and history. Some cadets demonstrated a fierce loyalty to the importance of managers, an importance we do not wish, but occasionally appear to, derogate. Some say they have learned enough about leaders (and themselves) to want careers as managers.

As we look to the future for this approach, we are reminded that many things remain to be done. We need more brief case biographies of military figures. We need films that focus on the aspects of appearance, delivery of speech and mannerisms that distinguish leaders. We need contrasting biographies of great managers such as Alfred P. Sloan and Robert MacNamara. But most of all, what remains to be done is to repair the imbalance in approaches to leadership. We do well in training for management; we need to do better in fostering a sensitivity to and an appreciation for leaders and an eventual development of the individual's potential for leadership.

IS FLYING ENOUGH? PILOT'S INTERPERSONAL SKILLS TRAINING

Judith Richter, Major, Israeli Air Force
Boston University, Dept of Sociology, Boston, Mass 02215

Abstract

Presented in this paper is the foremost part of the project of "management-commandship development and training" conducted in the Israeli Air Force. This section of the project stemmed from the recognition of a need to broaden the conception of skills necessary for pilot's flight activities, to also include commanding skills and the appropriate training program. What follows is an analysis of the nature of commandship during flight and the factors influencing the psychological aspects of this performance, as well as the functioning of pilots in senior command positions on the ground. We will also discuss some characteristics of the training program designed to train pilots in dealing more effectively with the managerial-commanding requirements of their work.

Pilots as commanders in flight

The present approach takes into consideration the inherent combination of flying and command duties and focuses on the latter. We differentiate between two main groups of factors which play a relevant role through the psychological aspects of pilot's performance: A. "Situational factors" - derived from the flight regulations and procedures. B. "personal factors" derived from the nature of the functions performed by the formation leader in flight, and from the follower's psychological reactions towards them.

A. Situational factors

1. During flight

a) Identical and simultaneous professional technical performance. During a flight mission, the leader and the follower perform identical professional tasks, however, the leader has the additional responsibility of command. This means that while actually flying he must exhibit also airmanship, to be aware of his follower's performance and occasionally to instruct him.

b) Continuous exposure of the leader's flying performance: This exposure greatly pressures the leader and according to pilots' reports, it affects his willingness to try unfamiliar and innovative maneuvers, in which a probable imperfect performance may degrade his professional image. Furthermore, this may result in the well known phenomenon of the leader deflecting the flight towards maneuvers where his best skills are manifested and excluding those parts of the flight about which he is less confident and more reluctant to expose.

c) Competition between leader and follower: A training method frequently used is built on rivalry between the "couple". The simultaneous performance of leader and follower, coupled with the continuous exposure of the leader, potentiates implicit competition between them. This competition is further encouraged by the realization that at the subsequent de-briefing session comparative evaluation of their performance will be conducted.

d) Interdependence: Aside from the rivalry, the flight regulations define generally a state of interdependence between the leader and the

follower. These regulations are, however, sufficiently flexible to allow for the leader's own personal commanding style to be revealed, and thus the extent of interdependence will be individually determined by the style of the leader.

e) Assessment of the leader: A common belief held among pilots is that the effectiveness of a couple depends, among other things, upon the interaction between the leader and the follower (that is, when flying as a unit against another formation). Yet, unlike the criteria for assessing the leader's technical achievements, those for assessing his commanding performance are far from being clear-cut. Hence the leader may see it profitable to display his technical flying skills at the expense of neglecting his commanding duties in order to achieve a better defined approval.

2. On ground

a) Inherent mutual evaluation: During the immediate de-briefing session, the leader and the follower offer mutual criticism. Naturally, they may both find it a strained situation, the extent of which depends upon their interpersonal relations.

b) Constant exposure of leader's overall flying performance: Leader's flying ability gains further attention by regular visual display of his flying achievements, constantly displayed on a board throughout several months of the training period. According to pilots the desire to display oneself favorably on the board has a profound influence on flight performance. There is a tendency to concentrate on those flight aspects that are later displayed at the expense of those that are not, even though the latter may be of equal importance.

c) Group pressure on the leader: During the group de-briefing session, taking place at the end of the day, the leader is exposed to critical evaluation of the entire squadron. The psychological pressure at this time is further increased since more objective evidence (films, recordings, etc) are inescapable. A discrepancy between the couple de-briefing and the squadron de-briefing can, at best, be explained as a misunderstanding or, at worst, as an indication of his poor airmanship.

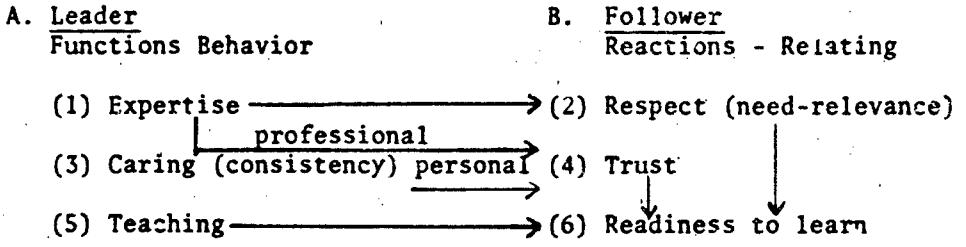
Some aspects of pilot's work as commander in flight that arise from the required combination of professional and commanding skills are presented above. As a result of this complexity it is possible that the leader will be stressed to emphasize his strong points while underplaying his weak ones. Hence, the follower is also put under pressure desiring to demonstrate his technical flying ability while at the same time trying to maintain a certain loyalty to his leader in not revealing the latter's weak points.

In general, this represents the struggle for professional authority. The leader, although having a formally defined commanding authority, still has to demonstrate certain capabilities in order to gain professional prestige, which in turn influence his performance as a pilot and a leader.

B. Personal factors.

The proceeding discussion and analysis suggest the existence of several factors that shape the relationship between the leader and the follower. These factors relate to leader's functions and behavior, affecting follower's reactions.

Leader's and follower's interactions:



Pilots are highly professionally trained and the main characteristic of a person in the squadron is his flying ability, insofar as his expertise (1) has been demonstrated. Being an expert is a major function of the leader determining his relationship with the follower, since only after expertise has been adequately manifested can the leader demand and acquire the follower's respect (2). It is quite important, however, that this expertise will be related to the follower's need and thus also of relevance to their joined work and command relationship. In addition, it is of great advantage that the follower will believe that his leader cares for him. This caring (3) function and behavior (may sometimes compensate even for lack of expertise) will lead the follower to trust (4) the leader. Another kind of trust may arise from the leader's expertise relying mainly on professional reasons while the one arising from the leader's behavior relies on personal ones. This issue of trust derives its unique significance from the "life and death" situation of combat pilots, thus the trust may relate to the physical level as well as to the emotional one. An important determinant for the development of trust is the consistency of the leader's caring manifestation.

The above are antecedents to leader's effective teaching (5) since follower's readiness to learn (6) is largely determined by the extent to which the leader is perceived as a flying expert and as trustful person. Leaders seldom differ in their ability and willingness to teach junior pilots during flight. It is up to the leader's style and decision as to what extent he might carry out this usually unpreferred role.

The Training Program

By analogy to a tennis match, while the follower is just one of the players, the leader is playing simultaneously the role of: the opponent, the coach, the referee and the press reporter evaluating the players.

After studying the existing training program, we noticed that it focused mainly on the acquisition of technical flying skills and that, in fact, the average air-crew man receives no training to help equip him for the command roles he would assume. The idea was to build a training program which would help pilots to cope with the different roles imposed on them while in flight, and to supply them the skills for interpersonal relations during flight and on ground. Our main contribution was in bringing to the leaders' awareness the conflict between their different roles and in trying to broaden their scope of interpersonal capabilities. In discussions with junior and senior pilots, we learned that the leader's style of behavior affects very clearly the follower's performance and thus it was this behavior that we addressed in our training.

More specifically we divided the difficulties brought up by the "situational" as well as the "personal" factors characterizing the pilot's work into two types: 1) the interpersonal difficulties - that relates to the leader's own integrity and play a role inside him to interfere with his ability to play the multiple role with due balance. 2) The projected difficulties - that are brought about by the follower's interpretation of the leader's action and interfere with the effectiveness of the relations between the leader and the follower.

In the "leaders' training program" we had three goals in mind that also set the general frame for this program's curriculum: a) The identification of potential interpersonal areas for leader's influence. b) Definition of a set of interpersonal skills that play a relevant role during flight. c) The development of training program to meet these requirements and its execution. As the teaching was based on the belief that training should respond to the specific problems brought up by the participants, the actual issues dealt with in each training session varied with the individuals comprising each group (fighter pilots, attack pilot, transportation pilots or system operators). In training we focused most frequently on the processes and communication styles in delivering instructions and gathering information while in the air. We also dealt with the modes of training the follower, the flexibility of the leader and his tolerance for the follower's mistakes, his willingness to share failure and success and his reaction to feedback of the junior pilot in the crew. In relation to the ground briefing and de-briefing we referred to the content as well as to the style and process of communication. This allowed us to deal with the leader's influence on the mutual trust level as caused by his noticeable expressed care for the followers. We also dealt with the leader's way of handling unforeseen difficulties and his style in encouraging the follower's learning process as part of his teaching function's fulfillment.

The training program was set up in three day workshop sessions. The number of participants was limited. The learning was based on synthesized methods of cognitive and experiential learning. To accomplish this we use several techniques of case study, demonstrations, simulations, role playing and interpersonal feedback. Throughout the workshops we also concentrated on the formation of each pilot's "response pool" preparing him to his future roles.

Two other positions for which workshops were set up were the positions of squadron commander's deputies and the squadron commander himself. The workshop for the deputies was based on broadening the concept of their position to include management roles. In the workshop we analyzed several social aspects of the pilot's environment change as he is promoted to this command position. We tried to enhance the participants' capabilities in situation sensitivity, awareness to their own management styles and the impact of their behavior on others. Since the participants originated in different squadrons we could make situation comparisons to facilitate evaluation and improvement of participants' effectiveness.

In squadron commanders' training we focused on the skills required for decision making. Three separate workshops were designed to deal with: a) The personal style in decision making. b) Chairing meetings and decision making groups. c) Management by objectives.

All the workshops were carried out in a non-military setting. Most of the participants knew each other from the past and few shared flight experiences. We tried to encourage the participants to make use of this information in order to improve their feedback and to adapt it to each participant's squadron and mission nature. One of the interesting points of the whole project was that participants in one workshop were very likely to meet in the next stage and personal change and/or improvement could be observed and discussed.

Leader Effectiveness as a Function of Leader's Sex and Leadership Style

Mickey R. Dansby
Department of Behavioral Sciences and Leadership
USAF Academy, CO 80840

Abstract

An experiment was conducted in which leader effectiveness was observed as a function of leader's sex and leadership style. Although leadership style was a significant predictor of leader effectiveness, neither leader sex nor the sex X style interaction significantly affected leader effectiveness. Some evidence of sex stereotyping was observed. The results were interpreted as supporting Fiedler's contingency model of leader effectiveness.

Many social scientists have reported various sexual stereotypes which are common to western society. Unfortunately, little research has investigated how these stereotypes may influence the effectiveness of leaders of each sex differently. Most research on sex stereotypes in leadership has concentrated either on follower impressions and attitudes or on cataloging differences in behavior between male and female leaders. Thus, more research is needed which investigates whether the leader's sex, either alone or in concert with other variables, has a significant impact on his/her effectiveness.

Contemporary theories of leadership are, for the most part, interactional in nature. Leadership is conceived as a social influence process involving the interaction of characteristics of the leader, the followers, and the situation. One such theory is the contingency theory of leadership proposed by Fiedler (1967). Fiedler proposes that the leader's effectiveness (i.e., group productivity) is contingent upon his/her style of leadership in interaction with the favorability of the situation for the leader. Two major styles of leadership, task-oriented (structuring) and human relations-oriented (considerate), are considered in Fiedler's theory. In the present work, we sought to understand how the leader's sex and stereotypes about what may be appropriate sex role behaviors for leaders may influence the interaction between leadership style and situational favorability in determining group performance for small, mixed-sex groups.

Method

After reviewing relevant research and considering the propositions of Fiedler's theory, as well as the proposal by Parsons and Bales (1955) that male leaders are expected to be instrumental (structuring) while female leaders are expected to be expressive (considerate), the author designed an experiment in which leader effectiveness could be observed as a function of the leader's sex and leadership style. A randomized blocks factorial design was employed in which leader sex (male, female), leadership style (structuring, considerate), and task order (TV task first, word task first) were the independent variables. The blocking factor (time of experimental participation) did not affect the results significantly, and it was subsequently dropped from the model. Dependent variables were 31 interval measures and 22 frequency measures of the groups' performances on the tasks, task strategies, and impressions.

PRECEDING PAGE BLANK-NOT FILMED

Preliminary testing had indicated that the tasks, in terms of Fiedler's contingency theory, were structured, the leader/member relations were good, and the leader position power was weak. Final ratings by judges not involved in the experiment confirmed these preliminary results for one task, a television game in which the players "hit" an electronic ball back and forth with electronic paddles for as long as possible. For the other task, a word game in which group members made as many words as possible from a given configuration of letters, the preliminary ratings of leader/member relations and leader power were confirmed, but the task failed to reach the criterion established by Fiedler for a task to be considered structured. Therefore, for the TV task, the favorability for the leader was in the high range according to Fiedler's theory; for the word task, the favorability fell into the moderate range. Fiedler's theory makes divergent predictions as to which leadership style will be most effective in these two favorability ranges. In the former range, Fiedler says that task-oriented (structuring) leaders are more effective; in the latter range, human relations-oriented (considerate) leaders are said to be more effective.

Besides having different structure levels, the two tasks were rated by judges as different in appropriateness for male and female participants. While the word task was rated equally appropriate for males and females, the TV task was seen as favoring males.

Based on the author's preliminary assumption that the favorability for the leader would be in the high range, and upon previous research and theory, several hypotheses concerning the effects of leader sex and leadership style were posited. These hypotheses were divided into three classes, depending on the type of outcome measures observed (i.e., performance measures, participant ratings, and strategies). The author proposed that where the favorability for the leader is high, the following effects on group performance will be observed:

(1) Structure-oriented leaders will be more effective than consideration-oriented leaders;

(2) Male leaders will be more effective than female leaders; and

(3) These main effects will be tempered by an interaction between leader sex and leadership style, with male leaders being more effective using a structuring style and both sexes being equally effective using a considerate style.

For the participant rating measures, the following hypotheses were proposed:

(4) The leader's sex will affect group members' ratings of group interaction and of the leader's behavior. Furthermore, male and female leaders will differ in their perceptions of group interaction;

(5) Structuring leaders will be rated by group members as more structuring than considerate leaders, and vice-versa;

(6) Leader sex and leadership style will have an interactive effect upon group members' ratings of the leader and the situation; and

(7) Female followers will rate their leaders more considerate, when compared to male followers.

One main hypothesis concerning strategy was posited:

(8) Structuring leaders will be more concerned with organization and efficiency in orienting toward the task goal, while considerate leaders

will be more concerned with involving all group members in the task and allowing each person to contribute to the task strategy.

Finally, a hypothesis concerning Fiedler's measure of leadership orientation, the Least Preferred Coworker Scale (LPC), was offered:

(9) Males will, in general, be more task-oriented than females.

Briefly, the procedures for the experiment were as follows:

Groups composed of two males and two females were formed from 256 undergraduate volunteers solicited via sign-up sheets and participating in partial fulfillment of a course requirement. These 64 groups were assigned randomly to each of the eight combinations of independent variables. The leader of each group, though ostensibly selected via a "leadership test," was selected at random according to appropriate sex.

As each group arrived at the group dynamics laboratory, they were briefed on what to expect during the experiment. Then, the experimenter informed the group that they would have a chance to win a $20 prize in the experiment, with their chances of winning the prize increasing as a function of good group performance on the tasks. The leader was "selected" via the "leadership test." While the experimenter was "scoring" the test, each group member completed the LPC questionnaire. The experimenter then announced the leader and escorted the leader into another room to receive instructions. While the experimenter gave these instructions, the other group members participated in a group discussion of a personal relations problem. During the briefing, the experimenter instructed the leader in the use of the appropriate leadership style and explained the rules for the first task.

The leader and experimenter then returned to the group, where the leader was afforded a brief period to prepare the group for the first task. The experimenter observed the group's preparations and their subsequent task performance from an inconspicuous vantage point. After the group had completed the first task, the experimenter escorted the leader to the private briefing room and gave the leader instructions for the second task. The leadership style briefing was not repeated, nor was the leader given feedback as to how well the group had performed.

Upon returning, the leader once more was allowed to prepare the group for the task. When the second task was over, the experimenter administered a questionnaire containing various rating scales designed to assess the group members' reactions to the leader and the situation. When all members had completed the questionnaire, the experimenter explained the purposes of the research and thanked the group for their participation.

Results

Results of the experiment indicated that the manipulation of leadership style had been effective. Participant ratings of the leader behaviors and independent observations by other persons assigned to count leader behaviors agreed in indicating that structuring leaders were more structuring than considerate leaders and vice versa for considerate behaviors.

Effects of the independent variables were assessed using MANOVA for the 31 interval measures. The 22 frequency scores were transformed using an angular transformation and these scores were analyzed via ANOVA. All hypotheses were evaluated with an alpha of .05. Since the frequency scores could not be subjected bo MANOVA, results on these measures were examined in conjunction with significant multivariate effects for the interval

measures. Where specific a priori predictions had been made, treatment effects were evaluated using a per comparison error rate.

The measures of group performance (leader effectiveness) showed an overall effect for leadership style ($F_{3,54} = 3.92$), with score on the TV task being the only significant univariate effect ($F_{1,56} = 8.85$). Groups with structuring leaders performed better on this task than groups with considerate leaders. The predictions that male-led groups would be more effective and that there would be an interaction between leader sex and leadership style for performance scores were not supported.

On the rating measures, there were significant multivariate effects for leadership style ($F_{23,34} = 2.57$) and task order ($F_{23,34} = 2.36$). For structuring leaders, group atmosphere was rated worse ($F_{1,56} = 5.17$), leader structure was rated greater ($F_{1,56} = 7.90$), leader consideration was rated lower ($F_{1,56} = 22.03$), individual group member's performance was rated worse for the first task ($F_{1,56} = 5.61$), leader power was rated higher ($F_{1,56} = 16.71$), and group members' enjoyment of the experiment was rated lower than for considerate leaders ($F_{1,56} = 4.00$). Univariate task order effects were for the group members' ratings of task dimensions (structure, $F_{1,56} = 6.30$; solution multiplicity, $F_{1,56} = 15.82$), performance (by the individual, $F_{1,56} = 7.17$; and the group, $F_{1,56} = 12.35$), and attribution of individual responsibility for group performance ($F_{1,56} = 6.22$).

The predicted overall effects for leader sex and the interaction of leader sex and leadership style were not observed on the rating scores. However, several a priori hypotheses concerning rating scores were supported: (1) Male leaders were rated as more influential than female leaders, t (56) = 1.69; (2) male leaders rated the group situation less favorably, t (56) = 2.88; (3) male leaders were more critical of their groups' performances, t (56) = 2.28; and (4) female followers rated their leaders more considerate than did male followers, t (112) = 1.67.

On the strategy measures, there was only one multivariate effect, leadership style ($F_{5,52} = 5.09$). Groups with structuring leaders used fewer players on the TV task ($F_{1,56} = 14.52$), more word listers on the word task ($F_{1,56} = 4.69$), and more people to check words in the dictionary ($F_{1,56} = 4.34$) when compared to groups with considerate leaders. Analysis of frequency data indicated that structuring leaders, as contrasted to considerate leaders, appointed more group members to specific positions ($t_{inf.} = 6.48$), used fewer female players on the TV task ($F_{1,inf.} = 15.97$), and appointed more male word listers ($F_{1,inf.} = 7.68$). Furthermore, groups with structuring leaders were more likely ($F_{1,inf.} = 4.38$) to use a specific cooperative strategy on the TV game, and chose a strategy less frequently on the word task where all group members called words out to a

single word lister ($F_{1,inf.} = 6.43$). When compared across all groups, males participated significantly more often than females in skilled positions on the TV task ($t_{56} = 6.21$), while the reverse was true for the word task ($t_{56} = 2.67$).

Males and females failed to show the predicted difference in leadership orientation, as measured by Fiedler's LPC Scale. Furthermore, leader's LPC score was not a significant predictor of group performance.

Discussion

The results were interpreted as indicating that leadership style is an important determinant of leader effectiveness, but that the leader's sex and the interaction of leader sex and leadership style may not be all that important in determining leader effectiveness. Generalization from the present situation to others should be approached with caution, however, since college students in an academic setting may show different responses to female leaders than people in general, and since leadership was observed in a relatively restricted group/task situation. More research is needed to see if the present findings are replicated in other situations.

The differences in rating scores support a moderate to weak effect for leader sex, but not for the interaction of leader sex and leadership style. Perhaps the effects of sex-stereotyping are diminishing in the wake of current social movements. We should continue to observe the course of sex stereotypes as more women come to hold responsible positions in our society.

The findings are interpreted as supporting Fiedler's contingency theory of leadership. In the highly favorable situation for the leader (when the group was working on the TV task), groups led by structuring leaders performed better. In the moderately favorable situation, we saw a reversal in which style was more effective, though style was not a significant predictor of success on the word task. These results are generally in agreement with what we would expect based on Fiedler's theory.

Since the present research failed to find the pervasive effects of sex stereotyping reported by other authors and as predicted by Parsons and Bales' (1955), it was suggested that further research be conducted to find out where and when sex stereotypes may be important determinants of leadership effectiveness and group members' impressions.

References

Fiedler, F. E. A theory of leadership effectiveness. New York: McGraw-Hill, 1967.

Parsons, T. & Bates, R. Family socialization and interaction process. Glencoe, Ill.: Free Press, 1955.

Gender and Organizational Leadership: A Comparison of Two Models[1]

H. Wayne Smith
Applied Science Associates, Inc.
Ft. Bliss, Texas 79916

Abstract

An organizational study was performed to integrate three lines of research (sex role stereotyping, attribution theory, and equity theory) which have provided some consistency to the investigation of gender and leadership. Male and female college students (N=120) participated in an in-basket simulation involving a "computerized" aeronautical engineering organization, into which personality and behavioral profiles of government workers purportedly were programmed. Multivariate analysis of behaviors and attitudes produced a model of gender effects on leadership. The model--focusing on performance, role conflict, reward allocation, and task assignment--was compared to a previous model based on those studies in the literature which had addressed either stereotyping, equity, or attribution in relation to gender and leadership.

Although an understanding of the relationship between gender and organizational leadership potentially is of great importance to the armed forces, little systematic work addressing pertinent underlying psychological processes has been accomplished. A promising approach is to integrate three relevant lines of research in social psychological theory: sex role stereotyping, attribution, and equity (Terborg & Ilgen, 1975). Analysis of available literature addressing these concepts gave rise to a general systems model (Figure 1) of attitudes toward women in male-dominated organizations.

A Systems Model

Stereotypic perceptions of women (Schein 1973, 1975) establish expectations of performance level. When this level is not exceeded by the woman, the observer's stereotypic view is reinforced. This view also is reinforced when the expected level is exceeded: attributions are made to such factors as luck, low task difficulty, and high effort level (Taynor & Deaux, 1973).

Perception of relative input to the organization and, consequently, allocation of organizational rewards also is influenced by attributional factors (Leventhal and Michaels, 1969). Thus, when success by a woman is attributed to relatively great effort when compared to that of a man, she receives proportionately larger organizational rewards. When success by a female is attributed to luck, however, the woman may expect to receive proportionately smaller rewards than a man. Success by males generally is attributed to skill, success by females to luck (Deaux & Emswiller, 1974). Males thus typically receive proportionately greater rewards. Another factor which may cause women to receive relatively small rewards is their alleged lack of managerial attributes (Schein 1973, 1975). Women are

[1]This paper is based on a portion of the author's doctoral dissertation, written at Texas Tech University under the direction of Clay E. George.

PRECEDING PAGE BLANK-NOT FILMED

expected to be less capable than men and thus receive fewer rewards and promotions than males. This maintenance of the low pay/low status position of females reinforces the stereotype of the relatively incompetent woman.

Stereotyping also influences both the internal and external involuntary constraints (Leventhal & Michaels, 1969, 1971) which restrict women in organizations (O'Leary, 1974) and cause them to suppress the exhibition of managerial job attributes (Schein, 1973). This suppression leads to poor performance (when compared to men) and thus to lower pay and fewer promotions. Conversely, high organizational input by women is attributed to the great effort required to overcome the involuntary constraints, and results in greater rewards than those allocated to equally performing males.

The Integration Study

Method. A study was conducted specifically to test these ideas. College students (N=120) served as subjects. Each was assigned an executive position in a simulated aeronautical research organization and completed 24 management-related in-basket problems. Subjects were told that: (1) the personality attributes and behavior patterns of a sample of actual government workers had been programmed into a computer; and (2) their managerial actions were being programmed as they completed a series of computer cards outlining actions to be taken in the in-basket problems. The deception was enhanced by the introduction of "computer noises".

Spurious feedback was provided by printouts indicating the organization's performance level as well as that of two equally qualified, equally ranked chief assistants--one male, the other female. On each problem, the subject had the option of delegation to a designated subordinate or of directing his/her personal attention to the problem. There were three feedback conditions: (1) equal performance; (2) superior performance by the male subordinate; and (3) superior performance by the female. Organizational performance was kept constant.

Subjects completed Likert-type scales to measure skill attribution and role comfort, a forced-choice measure of promotion preference, a measure of merit raise allocation, and the Work Scale of the Job Descriptive Index. The primary analysis was a 2X3 MANOVA. Sex of subject and feedback condition were the treatment variables. Other analyses included univariate assessments of attribution, promotion, and problem delegation responses.

Results. There was a MANOVA main effect for feedback [$F(14,216) = 17.50$, $p > .0001$]. High performing subordinates, regardless of sex, were given higher skill attributions than poorer performing subordinates [$F(2,114) = 110.84$, $p > .0001$], with no indication of differential skill attribution for equal performance by males and females. Under equal performance--and counter to the involuntary constraint hypothesis of equity theory--male subjects awarded more promotions to male subordinates than to female subordinates [$X^2(1) = 5.00$, $p < .03$]. The amount of merit raise awarded was a function of perceived organizational input, supporting equity theory, [$F(2,114) = 124.40$, $p < .0001$].

A panel of experts judged the in-basket problems on their relative difficulty and importance. Subjects of both sexes: (1) spent more time on difficult/important problems relative to trivial issues when the male subordinate was superior [$F(2,114) = 6.08$, $p < .01$]; (2) delegated more trivial problems to female subordinates than to males [$X^2(2) = 8.10$,

$p < .02$ for male subjects; $X^2(2) = 20.10$, $p < .01$ for female subjects]; and (3) delegated more difficult/important problems to males than to females [$X^2(2) = 10.0$, $p < .01$ for male subjects; $X^2(2) = 9.10$, $p < .02$ for female subjects].

A Revised Systems Model

The study produced a revised systems model (Figure 2). In this model, the concepts of involuntary constraints and attribute suppression have been eliminated, and those of stereotype confirmation and role appropriateness have been added. Further, organizational reward allocation has been differentiated between promotions and merit raises. The relative importance of problems addressed by the subordinate also has been taken into account. The following description of the revised model is based on successful performance by both male and female subordinates under male supervision. This condition, in the real world, may be the most likely arena for organizational conflict.

In the revised model, sex role stereotyping emerged as a key concept, the effects of which are apparent in subtle ways. Females are given relatively simple assignments, leaving the more important tasks for their male colleagues. This contributes to a tendency for male managers to promote other males over equally qualified females. Supervisors, observing that male subordinates are capable of making important organizational contributions by solving important problems, recognize these achievements by awarding promotions. However, females--relegated to relatively unimportant activities--are perceived as producing minor organizational contributions and receive fewer promotions than their male colleagues. Managers (regardless of sex) feel the necessity for personal involvement with all problems--even those which are relatively trivial--if the top-performing assistant is a woman. If the superior lieutenant is a man (i.e., the managerial stereotype is confirmed), the manager feels freedom to devote most of his/her attention to significant problems and to spend less time on peripheral issues.

Thus, at least three organizational factors contribute to stereotypic perceptions of females: 1) differential work assignment on the basis of the importance/difficulty of the problem; 2) promotion bias (in the case of male supervisors) in favor of males; and 3) the perceived necessity on the part of the supervisor for significant personal involvement in peripheral issues when the top-performing subordinate is female. These factors reinforce the prevalence of a relatively inferior position for females within the organization. Supervisors have less confidence in females than in males and tend to award them few promotions over males of equal ability. These factors are consistent with the view expressed by Schein (1973, 1975) that both male and female managers associate those attributes necessary for success in management with traits displayed by males in general.

Skill attribution is a function of performance, without regard to sex. Females who perform well receive acknowledgement of their skills rather than having their success attributed to luck. Even though males work at more difficult problems than those faced by females, males and females who are equally successful are awarded equal skill attributions. However, although there was no direct evidence of this, managers may nonetheless feel that females are less independent, objective, and competitive than men.

Thus, the general perceived appropriateness of males for the task situation (a result of stereotyping) may have greater impact on promotions and task assignment than skill recognition.

Although the perception of relative input to an organization is influenced by attributional factors, it is modified by stereotypic preconceptions. Success by a female will, for instance, result in supervisory recognition of her high skill level. Merit raise apportionment also can be tied directly to perceived performance. However, if the female is competing for promotion with an equally successful male, the managerial decision (at least among male supervisors) may be to promote the male on the basis of his allegedly superior managerial attributes. Although the female will receive token recognition, the male will be favored for promotion.

References

Deaux, K. & Emswiller, T. Explanation of successful performance on sex-linked tasks: What is skill for the male is luck for the female. Journal of Personality and Social Psychology, 1974, 29, 80-85.

Leventhal, G.S. & Michaels, J.W. Extending the equity model: Perception of inputs and allocation of reward as a function of duration and quality of performance. Journal of Personality and Social Psychology, 1969, 12, 303-309.

Leventhal, G.S. & Michaels, J.W. Locus of cause and equity motivation as determinants of reward allocation. Journal of Personality and Social Psychology, 1971, 17, 229-235.

O'Leary, V.E. Some attitudinal barriers to occupational aspirations in women. Psychological Bulletin, 1974, 81, 809-825.

Schein, V.E. The relationship between sex role stereotypes and requisite management characteristics. Journal of Applied Psychology, 1975, 57, 95-100.

Schein, V.E. The relationship between sex role stereotypes and requisite management characteristics among female managers. Journal of Applied Psychology, 1975, 60, 340-344.

Taynor, J. & Deaux, K. When women are more deserving than men: Equity, attribution, and perceived sex differences. Journal of Personality and Social Psychology, 1973, 28, 360-367.

Terborg, J.R. & Ilgen, D.R. A theoretical approach to sex discrimination in traditionally male occupations. Organizational Behavior and Human Performance, 1975, 13, 352-376.

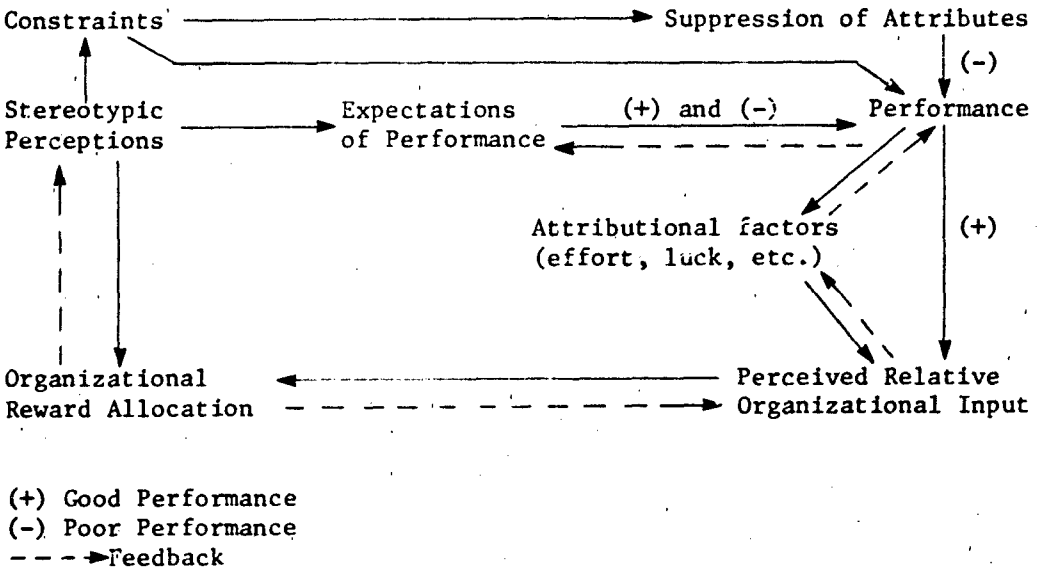


Figure 1. General Systems Model

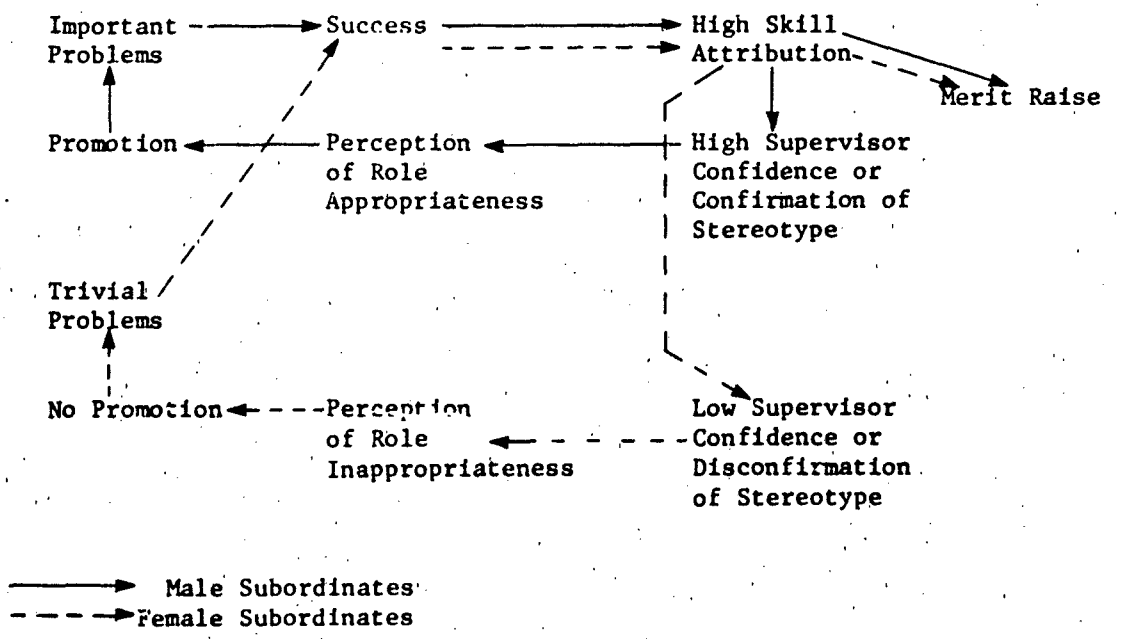


Figure 2. Revised Systems Model

Prediction of Military Order of Merit Using Assessment Center Methodology

Hal W. Hendrick
University of Southern California
Los Angeles, California 90007

Abstract

Using the United States Air Force Academy Group Reaction Course as a leadership assessment center, global ratings of leadership potential were determined for one half of an entering cadet class, using a five point scale. These assessments then were used to predict Military Order of Merit ratings at the end of the cadets' fourth class (freshman) year. The predictive validity of the assessment center ratings was found significant beyond the .01 level. Potential applications of similar assessments in cadet selection and training are discussed.

During the past two decades the management assessment center has emerged as the most effective means of assessing potential for promotion across a broad spectrum of organizations and settings. For example, from a review of over 50 studies conducted in different organizations Huck (1973) found that all showed positive findings. Byham (1970), based on a review of some 23 studies conducted in 20 companies operating management assessment centers, found that 22 showed assessment centers to be more effective than other approaches, and the remaining study showed the assessment center method to be as effective as some other approaches. Although correlations with various managerial and leadership performance criteria have occasionally exceeded .60, validity coefficients of between .30 and .50 have more typically been found for assessment center ratings.

The "original" assessment process on which current management centers are based was that developed for the selection of personnel for the Office of Strategic Services (OSS) during World War II. The predictive validity of the OSS assessments, using overseas staff appraisals as the criterion, was .37 for one OSS assessment center, and .53 for the other, corrected for restriction in range (OSS Assessment Staff, 1948). Wiggins (1973), using improved methods for estimating outcomes and predictions, found that 77% of the first OSS center's decisions, and 84% of the second center's judgements were correct.

At the core of the OSS assessment center approach were a series of leaderless group situational tests which were designed to assess the leadership and related skills of the participants. Typical of these situational tests was the Brook Test in which a group of six candidates were taken to an eight foot wide stream. On one bank was a heavy rock, on the other a log. A number of short boards were present on the side of the stream where the group stood. Also present were a barrel with both ends missing, a pully, and three lengths of rope. The candidates were told: "In this problem you have to use your imagination. Before you, you see a raging torrent so deep and so fast that it is quite impossible to rest anything upon the bottom of the stream. The banks are sheer, so it will be impossible to work except from the top of them. You are on a mission in the field, and having come to this

PRECEDING PAGE BLANK-NOT FILMED

brook, you are faced with the task of transporting this delicate range finder, skillfully camouflaged as a log, to the far bank, and of bringing back that box of percussion caps, camouflaged as a rock, to this side. In carrying out this assignment, you may make use of any materials around here. When the job is done, all of you, as well as any material you used, are to be back on this side" (OSS Assessment Staff, 1948, pp. 95-96). Following World War II these OSS and highly similar exercises were incorporated into various military training programs in the United States armed forces. A set of 12 such exercises were constructed at the United States Air Force Academy, where they became known as the "Group Reaction Course". The course primarily is used as part of the basic summar training program for entering cadets.

Given the success of the assessment center method in predicting future performance, it was hypothesized in the present study that global assessments of the leadership performance of new Air Force Academy Cadets on the Group Reaction Course would be predictive of their future leadership ratings by their peers, upper classmen, and Air Officer Commanding (tactical officer).

Method

Subjects

The subjects in this study were 423 entering cadets at the United States Air Force Academy. Assignment to the sample was based on being in the half of the entering class which was undergoing training on the Group Reaction Course at the time of the study. The sample group was found equivelent to the other half of the entering class in age, aptitude test scores, subsequent grade point averages, and distribution of ratings on various leadership criteria. The cadet assessee sample ranged in age from 17 to 24 years, with a modal age of 18. All assessees were male.

Apparatus

The apparatus for this study was the Group Reaction Course, located at the United States Air Force Academy, Colorado. The Group Reaction Course consisted of 12 constructed problem environments, similar to those developed and utilized in the OSS assessment centers (see OSS Assessment Staff, 1948 for detailed discriptions of these tasks). and a written set of instructions for each problem.

An assessment evaluation sheet was developed for assessor use in recording assessment ratings and comments. The sheet was divided into seven columns. The first column provided space for recording the names of six assessees; the right column provided space for entering written comments on each assessee's performance. The five narrow columns in between represented the five rating scale categories, and were labeled as follows: 1, Poor; 2, Below Average; 3, Average; 4, Above Average; 5, Outstanding. Below the category labels the sheet was divided into six sections by heavy horizontal lines. Each section contained six lines for recording the performance rating on a given assessee on each of six of the Group Reaction Course problems.

Procedure

24 first class cadets (seniors) at the United States Air Force Academy served as the assessors during the study. The assessors were divided into 12 two man teams. The assessors were allowed to choose their partners themselves. During the week prior to conducting the assessments the assessors underwent a training program to become throughly familiar with each of the 12 Group Reaction Course problems, including performing the problems several times. This instruction also included practice at using the assessment rating sheets. All of the assessors had undergone extensive prior experience at observing and assessing leadership skills as a routine aspect of their military training and academic courses in leadership at the Air Force Academy.

All of the assessments were completed during three four hour blocks, two on one day and the third on the following morning. Approximately one-third of the assessees were assigned to each block of time. When the assessees arrived at the Group Reaction Course they were divided into groups of six, and two groups were assigned to each pair of assessors. Group composition was determined by having the assessees "fall in" to a military formation of rows and columns, and then dividing each row up into sets of six. In order to accomodate the entire cadet sample, three of the groups were assigned an additional assessee; thus, there were 67 groups of six and 3 groups of seven for a total of 70 groups of assessees.

Each pair of assessee groups was assigned to one of the 12 Group Reaction Course tasks. The assessors then designated one of the two groups to perform the task, and the other to observe. Upon the signal to begin from the Group Reaction Course Officer, one of the two assessors read the instructions for the assigned task to the group and then directed them to begin. When all of the assigned groups either had finished the task or run out of the allotted time, the Group Reaction Course Officer signaled the groups to proceed to the next task, and all groups rotated in a clockwise direction. On the second task, the second group assigned to each assessor pair was designated to perform the exercise, while the first group rested and observed. This alternating procedure was repeated until each group had performed six of the tasks, and observed their paired group perform the other six exercises.

During the performance of each task the two assessors each observed the performance of all six group members. Immediately following each task the two assessors independently rated each of the six assessees global leadership performance on the rating sheet, using the 5 point scale provided, and entering any pertinent comments on each assessees behavior. At the end of the four hour block the assessees were dismissed and the assessors independently determined their composite ratings for each of their 12 assesses. The two assessors assigned to each pair of groups then compared their ratings for each individual. In each case where their composite ratings of an assessee differed, the assessors compared their ratings of the assessee's performance on each task and discussed specific aspects of the assessee's performance until they reached a consensus composite rating. In approximately 80% of the cases the composite ratings of the two assessors were identical. In about 15 % of the cases the composite ratings differed by one, and in less than 5% there were differences of two or greater. In all cases, assessor differences

in their initial global ratings were resolved by consensus.

At the end of the assessees' first academic year at the Air Force Academy they each were rated on their leadership performance by their respective (a) Air Officer Commanding (tactical officer), (b) Squadron upper class cadets, and (c) peers who were their squadron class mates. These ratings for each cadet then were combined to provide a composite rating which was used to determine the cadet's leadership rank within his class, or Military Order of Merit (MOM). In order to determine the predictive validity of the Group Reaction Course leadership assessment ratings, an analysis of variance was conducted for each of the four sets of criterian scores.

Results

Table 1 shows the number of assessees who were assigned to each of the five Group Reaction Course assessment rating categories. The distribution approximates normality in its shape. Also depicted in Table 1 are the mean leadership ratings for those assessees falling into each assessment rating category, and their mean Military Order of Merit Rankings which are based on the composite leadership ratings. For each criteria, the mean leadership ratings vary systematically in the expected direction as a function of the group reaction course assessment ratings. Military Order of Merit mean ranks for the assessee group varied systematically from 572 for those cadets assigned an assessment rating of "1, Poor" to 352 for those assigned an assessment score of "5, Outstanding".

Table 1

Mean Cadet Leadership Ratings and Ranking for Each Group Reaction Course Assessment Rating Category

	Group Reaction Course assessment category				
Type of leadership rating	1 Poor n=21	2 Below Av. n=75	3 Average n=197	4 Above Av. n=106	5 Outstanding n=24
Peer	3162	3268	2965	2558	2104
Upper Class	3360	3285	3020	2742	2438
AOC	3258	3233	3024	2674	2321
Composite	3252	3145	2949	2723	2419
MOM Rank	572	566	513	441	352

Note. All leadership ratings were assigned 10 months after the assessments.

In order to determine if these systematic relationships between the Group Reaction Course assessments and the four leadership criteria could be expected to occur in the population represented by the assessees in this study, a series of analysis of variance computations were performed. These are summarized in Table 2. For all three separate criteria, as well as for the composite rating the F ratios were significant at beyond the .01 level. It was concluded that the Group Reaction Course assessment ratings are predictive of future cadet leadership performance as determined by the criteria used in this study.

Table 2

Summary of Analysis of Variance for the Various Leadership Ratings

Source of variation	*df*	*MS*	*F*
		Peer Ratings	
Between	4	9970532.63	3.47*
Within	418	2871778.47	
		Upper Class Ratings	
Between	4	5831568.74	4.51*
Within	418	1292020.12	
		Air Officer Commanding Ratings	
Between	4	6653921.36	3.94*
Within	418	1688772.15	
		Composite Ratings	
Between	4	6595254.89	3.99*
Within	418	1654128.86	

*$p < .01$.

Discussion

The results of this study are consistent with the literature on the validity of assessment center methodology, including the use of leaderless group simulation tasks to assess leadership skills.

From a utilitarian standpoint, these results suggest that collecting assessments on entering cadets as a routine part of their Group Reaction Course training could provide for early identification of those individuals needing supportive instruction to bolster their leadership skills. These results also suggest that the Group Reaction Course could be used as a part of the cadet selection process. The end result could be more effective applicant screening and an upgrading of the leadership potential of the cadet wing.

References

Huck, J. R. Assessment centers: A review of the external and internal validities. Personnel Psychology, 1973, 26, 191-212.

OSS Assessment Staff. Assessment of men. New York: Rinehart, 1948.

Perceptions of Psychological Influence: The Role of Person-Environment Fit in Subordinate Reactions to Leadership

Lawrence R. James and S. B. Sells
Institute of Behavioral Research
Texas Christian University
Fort Worth, Texas 76129

Abstract

A hypothesis of consistency in cognitive information processing of psychological influence perceptions (perceived influence on decisions made by a supervisor) was proposed and tested. The hypothesis stated that if psychological influence was important to a subordinate, then he/she would (a) selectively attend to supervisor behaviors that reflected influence opportunities and (b) employ psychological influence perceptions in behavioral decisions (performance) and affective reactions (anxiety, satisfaction). The hypothesis was supported in a study of 362 Navy enlisted aircraft maintenance personnel, where the salience of psychological influence was determined by assessing the fit between personal characteristics of a subordinate (e.g., fear of failure) and the degree of overload in the work environment.

The objective of this study was to effect a coupling between (a) selective attention as it relates to subordinates' perceptions of their influence on decisions made by their supervisors and (b) relationships between such subordinates' influence perceptions and their behavior/affect. The term "psychological influence" is employed to refer to subordinates' perceptions of their influence on supervisors' decisions (James, Gent, Hater, & Coray, 1979). In effect, the question is whether psychological influence is important to a given type of individual in a given type of context. If psychological influence is important to a given type of individual, then he/she should selectively attend to behaviors on the part of the supervisor that reflect influence opportunities (e.g., participation in decision making) and control (e.g., close supervision). Furthermore, it seems reasonable to expect that if the need for influence is of sufficient importance to effect selective attention, then the resulting psychological influence perceptions would be employed to make decisions regarding behavior (e.g., job performance) and to provide information input for affective reactions (e.g., anxiety, job satisfaction).

Identification of variables that predispose subordinates to be selectively attentive to influence opportunities and control proceeded by examining the environmental contexts to be studied for salient environmental presses (Bronfenbrenner, 1977). The study sample consisted of 362 enlisted aircraft maintenance personnel in Navy Air Training Commands. Interviews with representatives of the commands helped to identify an important environmental press, namely the extent to which the maintenance personnel were "overloaded" by such things as extended work hours, undermanning, and pressures from sources above the supervisor for high productivity. The next step in the process was to identify person variables that interact with

PRECEDING PAGE BLANK-NOT FILMED

overload and predispose subordinates to value influence and thus to be selectively attentive to influence opportunities and control. The interviews described above, together with reviews of the overload, personality, and influence literatures, and some extrapolations, led to the selection of four interrelated person variables that presumably served this purpose. These variables were achievement motivation, self-esteem, need for certainty, and impulsiveness. The variables were employed to develop a person typology. Type I individuals were characterized by comparatively high levels of achievement motivation, self-esteem, and need for certainty, and low levels of impulsiveness. Type II individuals were described as having comparatively low levels of achievement motivation, self-esteem, and need for certainty, and high levels of impulsiveness.

Ratings of overload for each workgroup were provided by workgroup supervisors (n=107). A split at the mean overload score resulted in the classification of 57 workgroups as working in high overload conditions and 50 workgroups as working in low overload conditions. High versus low overload was then crossed with Type I versus Type II person types to provide four subgroups of subordinates. Within each subgroup, hypotheses were developed regarding expected relationships between subordinates' perceptions of psychological influence and (a) supervisors' use of influence opportunities and control (as described by a supervisor for each subordinate), and (b) subordinates' job performance, satisfaction with the Navy, and anxiety. The former set of relationships provided a basis for testing a selective attention hypothesis, while the latter relationships were employed to test a consistency hypothesis, as described below.

High Overload - Type I Subordinates. High overload conditions connote high environmental press for achievement, which should stimulate Type I's to manifest high levels of achievement motivation, self-esteem, certainty, and low impulsiveness (as long as successful accomplishment of work objectives is not severely restricted). Psychological influence should be salient to these individuals because influence provides occasions to clarify paths to goals, to share in responsibilities for outcomes, to demonstrate competence, and to attempt to insure that courses of action are planned and carried out in an organized manner. Thus, it is hypothesized that Type I subordinates in high overload conditions will be selectively attentive to supervisor-initiated influence opportunities, as evidenced by a significant positive relationship between supervisors' influence opportunities and subordinates' psychological influence. On the other hand, supervisors' use of control should be related negatively to subordinates' psychological influence inasmuch as control detracts from opportunities to affect decisions and may be regarded as questioning competence. Finally, given the assumed salience of psychological influence in this subgroup, positive psychological influence-performance/satisfaction and negative psychological influence-anxiety relationships are anticipated. That is, psychological influence is predicted to be consistently salient throughout cognitive information processing.

High Overload - Type II. The motivation to avoid failure and the low self-confidence thought to characterize Type II subordinates should be manifested in the form of a threat of failure in high overload conditions. Threat of failure should in turn lead to defensive actions to protect self-esteem; in such cases, defensive lack of effort is a strong possibility (cf.

Jones, 1973). From a cognitive standpoint, it is suggested that defensive lack of effort is in part a form of "psychological withdrawal." Psychological withdrawal involves cognitive self-removal from threatening, frustrating, and anxiety-producing situations, and is reflected by apathy, especially toward evaluative events. In the present circumstances, psychological withdrawal was expected to include (a) apathy toward perceptions of psychological influence and performance, because failure is viewed as likely and self-evaluations of responsibility for failure are avoided; (b) apathy toward supervisors' use of influence opportunities and control because these behaviors are likely to reflect negative evaluations of failure; and (c) a conscious effort to avoid having satisfaction and anxiety contingent on perceived events which reflect evaluations of failure. In short, psychological withdrawal implies that Type II subordinates in high overload conditions value neither performance nor psychological influence, and that no attempt will be made to form cognitive contingencies between evaluative environmental events (influence opportunities, control) and psychological influence, or between psychological influence and both performance and affect. Nonsignificant relationships were predicted, therefore, for all relationships with psychological influence (i.e., psychological influence is consistently non-salient).

Low Overload - Type I. Low overload suggests low environmental press for achievement, which may not stimulate Type I's to manifest high levels of achievement motivation, self-esteem, and so forth. Nevertheless, it is expected that Type I subordinates place value on psychological influence and will be selectively attentive to influence opportunities because such opportunities are intrinsically interesting in their own right, and further, provide a basis for increasing what may be low job challenge. Moreover, Type I's might value psychological influence and be attentive to influence opportunities because influence allows them to ensure that actions are planned and organized. Thus, positive influence opportunity-psychological influence relationships are predicted. Negative control-psychological influence relationships are also predicted, for the same reasons discussed for Type I's in high overload conditions. Finally, assuming that psychological influence perceptions are salient to Type I's in low overload conditions, it is anticipated that the perceptions will be related positively to performance and satisfaction, and negatively to anxiety.

Low Overload - Type II. The low environmental press for achievement implied by low overload is expected to suggest to Type II subordinates that there is a low probability of failure and a high probability for successful performance. Consequently, if it is assumed that (a) motivation to succeed will overcome motivation to avoid failure, (b) Type II's will seek self-enhancement in situations with high probability of success (cf. Jones, 1973), and (c) the motivation to succeed will dispel tendencies toward disorganization and impulsiveness, then it is possible to predict a "contingent" situation rather than the "psychological withdrawal" situation discussed for Type II's in high overload conditions.

In a contingent condition, Type II's should be concerned with performance because they regard successful performance as achievable. They should also value psychological influence and be attentive to supervisor behaviors because perceptions that one is even partially responsible for influencing successful outcomes is ego-reinforcing and self-enhancing. However, it is expected that the Type I subordinates will be attentive primarily to supervisors' use of control, rather than influence opportunities, because (a) the operationalization of influence opportunities includes provision of autonomy (i.e., encourages subordinate to act on own) and the setting of high objectives for the subordinate, and these supervisor behaviors could be threatening to Type II's, even in low overload conditions; and (b) the basic defensiveness of Type II's that evolves from fear of failure and protection of self-esteem might stimulate them to be sensitive primarily to decreases in negative feedback (i.e., decreases in control) rather than to increases in potentially threatening stimuli (i.e., increases in influence opportunities). Consequently, it is predicted that Type II's in low overload conditions will be selectively attentive to supervisors' use of control in the construction of psychological influence perceptions, where the relationship should be negative. Finally, given the expected salience of influence perceptions and performance, it is hypothesized that psychological influence will be related positively to performance and satisfaction, and negatively to anxiety.

Results and Discussion

Empirical tests of the hypotheses were predicated on (a) assessments of whether the relationships conformed to the predictions in each subgroup of subordinates, and (b) determinations of whether significant differences in relationships existed among the subgroups. Empirical confirmation of the predictions would support the consistency hypothesis. That is, for any subgroup, a significant relationship in column one or column two in Table 1 would suggest selective attention, and should be followed by significant psychological influence-performance/affect relationships in columns three through five in Table 1. In contrast, nonsignificant relationships in both columns one and two would imply lack of attentiveness, and thus perceptions of psychological influence should not be related significantly to the performance/affect variables.

Results of empirical analyses, presented in Table 1, include (a) unstandardized regression weights (b-weights) for the separate relationships of the supervisor behaviors and performance/affect variables with psychological influence in each subgroup, and (b) the results of planned comparison tests of parallelism for each set of relationships (i.e., across subgroups) with psychological influence. Comparison of the predicted direction and significance of relationships with the results in Table 1 demonstrates essentially perfect correspondence between a priori predictions and empirical results. Moreover, the results of the planned comparison tests of parallelism indicated that relationships between psychological influence and each of the leader behaviors and subordinate outcome criteria were moderated by subgroup membership, thus supporting the interaction hypothesis (although results for control did not meet conventional levels of significance.

Table 1

Unstandardized Regression Weights for Relationships Between Subordinates' Psychological Influence and Supervisors' Behaviors and Subordinates' Performance/Affect, and Results of Planned Comparison Tests of Parallelism

Subgroups	Relationships with Subordinates' Psychological Influence Perceptions				
	Influence Opportunities	Control	Performance	Satisfaction	Anxiety
High Overload-Type I (n = 98)	.45**	-.25*	.95**	.39**	-.53**
High Overload-Type II (n = 94)	-.02	-.19	.13	.07	-.13
Low Overload-Type I (n = 87)	.34**	-.56**	1.30**	.29**	-.32**
Low Overload-Type II (n = 83)	.09	-.47**	.73**	.29**	-.26**
F-ratio Based on Planned Comparison Test of Parallelism	9.54**	2.85[a]	4.31**	5.25*	4.64*

[a] $p < .09$ * $p < .05$ ** $p < .01$

In summary, the results supported the consistency hypothesis, suggesting that knowledge of the fit between persons and their environments provides a basis for predicting not only whether subordinates will be selectively attentive to leader behaviors that reflect influence opportunities, but also whether individuals will employ perceptions of influence in decision-making and attitude formation. These results suggest that effective leadership requires not only consideration of the degree of fit between subordinates and environments, but also the tailoring of leader behaviors to meet the needs of different types of subordinates.

References

Bronfenbrenner, U. Toward an experimental ecology of human development. American Psychologist, 1977, 32, 513-531.

James, L. R., Gent, M. J., Hater, J. J., & Coray, K. E. Correlates of psychological influence: An illustration of the psychological climate approach to work environment perceptions. Personnel Psychology, 1979, 563-588.

Jones, S. C. Self-and interpersonal evaluations: Esteem theories versus consistency theories. Psychological Bulletin, 1973, 79, 185-199.

Psychopharmacological Intervention in Bereavement and Grief

Victor M. Solomon
Headquarters, Air Reserve Personnel Center
Office of the Command Chaplain
Denver, Colorado 80280

Abstract

The use of psychoactive agents in grief therapy is examined from the perspectives of indication and contraindication. Sequelae of inadequate and inappropriate grief work may include a wide range of psychological, psychosomatic and social problems. The approach to chemotherapeutic intervention in grief therapy, herein developed, stresses that mourners should be given minor tranquilizers when adjunctive drug therapy is needed to help them cope with distressful symptoms. However, the normal grief process should not be interrupted with anti-depressants and other medications. The casual dispensing of drugs to bereaved persons in many health facilities may reflect more a requirement of hospital decorum than the mourner's needs, and may also relate to the physician's role of catering to oral needs during a regressive crisis experience. The need for research on the interface between psychopharmacology and bereavement has assumed critical dimensions in view of the epidemiological significance of grief.

Need for the Study

Students of human behavior have consistently observed that improper handling of grief can result in a variety of problems. Many emotional and psychosomatic disorders have their etiology in mismanaged grief experiences and improperly resolved mourning. Even the risk of suicide increases in the wake of bereavement. It was noted in one study that "certain persons, such as some widows of ages 40-60 who have been extremely dependent on their husbands, display exaggerated mourning and grief reactions associated with strong self-destructive urges."[1] Another research demonstrated a causal relationship between unresolved grief and juvenile delinquency.[2] Evidence was also discovered linking the death of a parent in childhood with depression "in a significant proportion of psychiatric patients."[3] Tens of millions of Americans are exposed to grief annually, and their ability to grieve appropriately may be a vital factor in determining their physical and mental health later in their lives.[4]

The bereavement experience is a traumatic upheaval for the person who has lost a loved one. The onset is helpless emotional disorganization. If nature is permitted to run its course without contrived interruptions, a reorganization and

reintegration of the individual's emotional life sets in. Unfortunately, our culture seems to work against the best interests of the mourner. He is insulated against the healing processes provided by nature, religion, and traditional institutional machinery of mourning subtly molded by ancient wisdom and immemorial usage, to help him face his loss in a constructive and healthy manner. Eric Lindemann postulated in his classic study of the Coconut Grove fire survivors that bereavement is a distinct syndrome with a variety of symptoms, many of which may be delayed for months and even years.

Ways must be found to recover and reexamine the traditional approach to the grief experience and to synthesize it with a realistic approach to psychopharmacology as a potentially useful adjunctive therapy in treating the bereaved. This can be a significant contribution to the general field of mental health.

Related Literature

Grief can no longer be ignored as a serious subject for medical scrutiny. Parkes writes: "Of all the functional mental disorders, the only one whose cause is known, whose symptomatology is stereotyped, and whose outcome is usually predictable is grief. That grief is a mental disorder there can be no doubt....The fact that it is usually transient and seldom treated by physicians is irrelevant to the issue.... The occurrence of depression as a pronounced symptom at some stage or other of most forms of grief has justified placing it among the reactive depressions."[5] G.L. Engel agrees that grief should be classified as a disease.[6] However, Schmale does not see grief as a "disease," and prefers to view it as "a natural psychological process...."[7]

Lindemann's research was a major step forward in the study of grief. A body of literature evolved around his pioneer work. It was an important stride beyond the intuitive notions of Freud, in Mourning and Melancholia[8] published in 1917. On Death And Dying by Elisabeth Kübler-Ross, published in 1969,[9] has influenced and dominated the swell-tide of interest in this field.

Psychopharmacology and Bereavement

Elisabeth Kübler-Ross inveighs against the "overuse of tranquilizing drugs....in the management of the unexpectedly bereaved...."[10] She contends that psychopharmacologic management of the bereaved is often more a social issue than a medical necessity. Physicians are especially threatened by the feelings of impotent rage displaced on them, and they use drugs "defensively." Drugs may interfere with mourning by removing the patient, as it were, from the confrontation with a tragic reality which eventually must be faced if it is to be

worked through. On the other hand, she does not entirely dismiss the value of drugs in treating a bereaved or dying person. Sedation may be indicated when requested by a mourner with a low threshhold of pain, and in the case of bystanders who are unable to tolerate the expression of anguish. As a rule, she insists that each case be individually evaluated, and that the needs of the bereaved be paramount in the decision to medicate.

Klerman, Kline and Danto caution against the routine medication of dying patients on similar philosophical grounds. Drugs are often more technology related than patient oriented, catering to rejection of confrontation with a meaningful experience. The best interests of the patient are frequently subordinated to the needs of the family and/or physician. Medication is viewed as "encapsulated empathy" administered in lieu of personal relationships.[11] Hollister, while acknowledging that "the bereaved state must be worked through psychologically," suggests that the symptoms of acute grief be treated with psychopharmacologic agents. For example, a few nights of treatment with sedative-hypnotic drugs may be helpful with insomnia; a brief period of assistance with sympathomimetic stimulants could be beneficial during the guilt stage of grief; and severe depression involving emotional withdrawal would require the administration of tricyclic anti-depressants.[12] Arkin agrees that "pharmacologic agents are most useful in the treatment of insomnia," but cautions against any interference with vital REM sleep, which is essential to the working through of grief. Therefore, hypnotic agents such as Doriden, Nolubar, and barbiturates must be avoided. Suppression of grief, through pharmacological agents, can eventuate in chronic pathological conditions.[13] Merliss prefers the use of anti-anxiety agents in the most common clinical states associated with the grief reaction. Although the symptoms often appear similar to those of depression, the manifestations can be understood primarily as aspects of an anxiety state. The normal grief reaction is, generally, self-limiting and morbidity can be attenuated by judicious use of supportive therapy and medication, preferably tranquilizers. Diazepam and chlorodiazepoxide, because of their broad spectrum effect, are the treatment of choice in normal grief reaction when supplemented by psychotherapy and social support. Peretz suggests a differential diagnosis to distinguish acute grief reactions from pathological depression. Psychotherapy and adjunctive medication are indicated when a transient grief reaction develops into morbid depression, "to make it possible for grief to proceed with favorable adaptive consequences."[14] In a sensitive

study of widows during the first year of bereavement, Glick, Weiss and Parkes found that "they were often helped by drugs that gave promise of preventing runaway anxieties, including tranquilizers, sleeping pills and alcohol," but felt ambivalent about medication as part of a general reaction to dependency.[15] Twycross, sensitive to the ambivalence in the scientific community concerning the relative efficacy of sympathetic listening and medication, suggests "that the total denial of psychopharmacologic support in bereavement is both shortsighted and inhumane," especially in a case where the bereaved are totally exhausted from the antecedent protracted care of a terminally ill loved one, and are without the "mental and physical reserves with which to face .... bereavement. The use of a bedtime hypnotic until, at least, after the funeral, would seem to be reasonable." Again, it is emphasized that when psychopharmacologic agents are prescribed, they should be used only as adjunct therapy and not as substitutes for human care.[16] Ferguson, in a 1964 survey of 86 widows, found that 67% took tranquilizers after the death of a husband, one-third for days or weeks, one-third for months and one-third for years. The two determining variables were the widow's personality (the dependent and dominant types were most vulnerable), and prior marital conflict. Drugs could help with coping, but they can also mask basic problems.

Maddison and Viola found little correlation between psychopharmacotherapy and subsequent improved health in a study of first year widows.[17] Wiener urges that the use of psychopharmacologic agents in grief reactions be limited to symptomatic treatment, and the total medical effort in assisting the patient to cope with bereavement.[18] Schmale insists that, except in attenuated grief situations, pharmacologic agents not be given to the bereaved. Pradham agrees, with the proviso that drugs may be needed in cases of severe and/or prolonged emotional disturbances.[19]

It is possible to generalize that, in ordinary grieving, psychopharmacologic agents should be avoided. Where the ego is threatened with serious disruption, drugs may be used with judicious restraint. Medication is most effective in the alleviation of secondary and tertiary symptoms when not in conflict with the "mourning process." Individuals who would, ordinarily, encounter difficulty in dealing with stress, and use drugs to help them cope, require pharmacologic help in facing the crisis of grief. Ivan K. Goldberg of Columbia University, noted psychopharmacologist and writer on bereavement, suggested to the researcher that diazepam (5mg., t.i.d.) is the treatment of choice for inhibited grief.

In conclusion, drugs are most useful when employed to "help the mourner mourn," i.e., to lower the inhibitions which interfere with appropriate mourning, and to enable him to free the affect.

Notes and References

1. R.E. Litman and N.L. Farberow, "Emergency Evaluation....," The Cry For Help, N.Y.: McGraw-Hill, 1961, p. 51.
2. M. Shoor and M.H. Speed, "Death, Delinquency ....," Psychiatric Quarterly, 37, 1965, pp. 540-58.
3. A.T. Beck, Depression, Phila., U. of Pa., 1975, pp. 540-58.
4. R. Fulton, "Death and the Self," J. of Religion and Health, 3,4, July 1964, pp. 359-68.
5. C.M. Parkes, "Bereavement and Mental Illness....," Brit. J. of Medical Psychology, 38, March 1965, pp. 1-12.
6. G.L. Engel, "Is Grief A Disease?" Psychcsomatic Medicine, 1961, 23:18.
7. A.Schmale, Jr., "Normal Grief Is Not A Disease," Psychopharmacological Agents for the Terminally Ill and Bereaved, I.K. Goldberg, et al, editors, New York: The Foundation of Thanatology/Columbia U. Press, 1972, p. 257.
8. S. Freud,"Mourning and Melancholia," The Standard Ed. of the Complete Psychological Works of Freud, London: Hogarth, 1957.
9. E. Kübler-Ross, On Death and Dying, N.Y.:Macmillan, 1969.
10. E. Kübler-Ross, "On the Use of Psychopharmacologic Agents....," Psychopharmacological Agents...., I.K. Goldberg, et al, editors, p. 3.
11. G.L. Klerman, "Drugs and Dying," (p. 15), N.S. Kline, "Indications for Psychopharmacologic Management....," (p.29), B.L. Danto, "The Cancer Patient....," (p. 39) Psychopharmacological Agents...., I.K. Goldberg, et al, editors.
12. L.E. Hollister, "Psychotherapeutic Drugs....," Psychopharmacological Agents...., I.K. Goldberg, et al, eds., p. 69.
13. A.M. Arkin, "The Use of Medications....," Psychopharmacological Agents...., I.K. Goldberg, et al, editors, p. 75.
14. D. Peretz, "Reaction to Loss," Loss and Grief:...., B. Schoenberg, et al, eds., N.Y., Columbia U., 1970, p. 28.
15. I.O. Glick, et al, The First Year of Bereavement, N.Y., John Wiley and Sons, 1974, p. 67.
16. R.G. Twycross, "Acute Grief:....," Acute Grief and the Funeral, V.R. Pine, et al, Springfield, Thomas, 1976, p.5,6.
17. D.C. Maddison and A. Viola, "The Health of Widows....," J. of Psychosomatic Research, 1968, 12:297.
18. A. Wiener, "The Use of Psychopharmacologic Agents....," Psychopharmacological Agents...., I.K. Goldberg, et al, p.255.
19. S.N. Pradham, "Basis for Psychopharmacotherapy of the Dying and the Bereaved," Psychopharmacological Agents...., I.K. Goldberg, et al, editors, p. 277.

The Impact of Stress in the Air Combat Environment

Dana R. Ideen, Capt, USAF
Jeffrey E. Kantor
Air Force Human Resources Laboratory
Brooks Air Force Base, Texas 78235

Abstract

A methodology was developed to measure the extent and impact of stress common to fighter missions flown in a combat environment. Data on stress, defined as "feelings of strain, pressure, or tension," was obtained from 563 members of the Red River Fighter Pilots Association. In addition, demographic, pilot experience, and an estimate of the number of sorties that could be flown in a 2-week combat surge effort were obtained. Demographic and pilot experience factors were found to be related to the stressfulness of mission events and the individual's sortie estimates. Implications for tactical fighter resource management are discussed.

Operations within an air combat environment are typically associated with subjective feelings of strain, pressure, and tension. Research and anecdotal reports indicate that these feelings are common to almost all personnel in combat and are typically not moderated, but rather increase with the individual's length of exposure to the combat setting (Shaffer, 1951). These subjective experiences may be subsumed under the rubric of psychological stress. Previous research on stress has led to one important relationship of particular relevance to performance in the air combat environment; i.e., the more complex the task, the greater the resultant impact from stress (Yerkes-Dodson Law). It is reasonable to assume that piloting a fighter to a target, avoiding the enemy threats, striking the target, and returning a possibly battle-damaged aircraft to base would be a very complex assignment of the type easily disrupted by stress. However, little research has been conducted on stress in the air combat environment and what has been done has concentrated on the physiological indicators of stress (e.g., Austin, 1969). This research was an attempt to assess psychological stress in the air combat environment with particular emphasis on identifying specific stressful events within the fighter aircraft mission. It should be noted that this is the first interim report of a proposed series in this area. This report documents the data that have been collected and some of the types of analyses possible with those data.

Method

Subjects

Subjects were 563 members of the Red River Valley Fighter Pilots' Association. Membership in this organization is limited to U.S. aircrews who have flown missions into heavily defended areas in North Vietnam. The members of this association are the U.S. pilots with the most recent combat experience.

Development of the Survey

To measure stress and its impact on air combat effectiveness, the Combat Stress Questionnaire was developed by Kantor, Klinestiver, and McFarlane (1978). The main body of the questionnaire is a listing of specific events of a fighter combat mission (e.g., briefing, takeoff, air-to-air encounters, etc.). For each event of the mission, the respondent is asked to indicate the frequency of occurrence and the stressfulness of that event. The questionnaire also requests information about the stressfulness of the various mission types, data on the pilot's flight experience, and estimates of the number of sorties that the respondent could fly in a 2-week surge period for each mission type.

Results and Discussion

Of the 563 respondents, the majority were either F-4 (n = 277) or F-105 (n = 152) pilots. The non F-4/F-105 pilots (n = 134) comprised a third group and included pilots of a variety of different aircraft including A-7, A-1, B-52, F-106, and others. Table 1 shows the top ten stressful events for all pilots compared to the top ten for F-4, F-105, and non F-4/F-105 pilots separately.

Table 1. Rankings of Stress Items by Aircraft Type[a]

	All Pilots			F-4 Pilots		F-105 Pilots		non F-4/F105 Pilots
Rank	Var	Name	Mean	Var	Mean	Var	Mean	Var	Mean
1.	29	SAMS, 80mm	3.84	29	3.88	29	3.91	29	3.70
2.	34	TOT-6 min	3.73	34	3.75	34	3.81	34	3.62
3.	35	A/C Dam-emer	3.60	21	3.73	35	3.72	35	3.44
4.	22	AA Tht-1+	3.57	22	3.71	30	3.68	30	3.41
5.	21	AA Tht-1	3.55	35	2.62	22	3.53	33	3.29
6.	30	80,57,37mm	3.51	30	3.45	21	3.51	22	3.28
7.	33	TOT-4 min	3.40	33	3.41	33	3.47	21	3.19
8.	26	A-G-6 passes	3.27	26	3.30	26	3.43	28	3.11
9.	32	TOT-2 min	3.17	27[b]	3.18	32	3.28	32	3.09
10.	28	A-G-1 pass	3.14	32	3.16	45[b]	3.28	26	2.99

[a] Survey scale: 1 - no stress; 2 - some stress; 3 - moderate stress; 4 - severe stress; 5 - very intense stress

[b] 27 - Air to ground: 3 passes
45 - Interdiction mission stress

Table 1 illustrates that there are few differences among th groups of pilots with respect to which events provoke the most st It is notable, however, that both the F-4 and the F-105 pilots' an stress ratings are consistently higher than the mean stress ratings from the non F-4/F-105 group, except for mission event 28, "Air to ground: 1 pass." Nearly all of the pilots had frequent combat missions requiring the "Air to ground: 1 pass" while the remaining events of Table 1 are primarily fighter mission factors. These data, then, support the assumption that pilots perceive events with consistency across aircraft types.

Table 2 lists all survey items where significant differences were found among the pilots of differing aircraft types. There are many significant differences between the non F-4/F-105 group and the fighter aircraft groups (F-4 and F-105). However, only three events were found to significantly differ between the F-4 and F-105: POD formation flying, the ground threat of 80, 57, and 37mm, and day instrument landings. Generally, Table 2 indicates that there are many areas of a combat mission that are significantly more stressful for fighter pilots than for the non F-4/F-105 pilots. Although event frequency of accomplishment data is currently not available, it is known that the fighter pilot groups flew more combat missions into the heavily defended areas of North Vietnam. Also, it is notable that the events where the non F-4/F-105 pilots are more stressed than the fighter pilots are non-combat; i.e., takeoffs and landings. This indicates that the normally stressful daily aircraft maneuvers tend to lose their potency as stress-provokers when included in the combat environment.

Table 2. Stress Items Showing Significant Differences Across Aircraft Type

Variable (no.)	Groups Compared (Mean[b])	p[a]
Day Takeoff, Clear (6)	non F-4/F-105(2.02) vs. F-4(1.79)	.002
Night Takeoff, Clear (8)	non F-4/F-105(2.61) vs. F-4(2.22)	.0001
Night Takeoff, Clear (8)	non F-4/F-105(2.61) vs. F-105(2.36)	.01
Night Takeoff, Inst (9)	non F-4/F-105(3.05) vs. F-4(2.78)	.008
Ground Aborts (10)	non F-4/F-105(1.80) vs. F-105(2.01)	.05
12-16 Ship Formation (15)	non F-4/F-105(2.02) vs. F-4(2.33)	.002
POD Formation (16)	non F-4/F-105(2.05) vs. F-4(2.62)	.00001
POD Formation (16)	F-4(2.62) vs. F-105(2.23)[c]	.0001
Air to air: 1 Encounter (21)	non F-4/F-105(3.19) vs. F-4(3.73)	.0001
Air to air: 1 Encounter (21)	non F-4/F-105(3.19) vs. F-105(3.51)	.021
Air to air: More than 1 Encounter (22)	non F-4/F-105(3.28) vs. F-4(3.71)	.006
Air to ground: 3 passes (27)	non F-4/F-105(2.89) vs. F-4(3.18)	.017
80,57,37mm (30)	non F-4/F-105(3.41) vs. F-105(3.68)	.013
80,57,37mm (30)	F-4(3.45) vs. F-105(3.68)[c]	.013
Landing, Day, Clear (38)	non F-4/F-105(1.71) vs. F-105(1.40)	.00001
Landing, Day, Clear (38)	non F-4/F-105(1.71) vs. F-4(1.42)	.00001
Landing, Day, Inst (39)	non F-4/F-105(2.21) vs. F-105(1.89)	.0003
Landing, Day, Inst (39)	F-4(2.07) vs. F-105(1.89)[c]	.014
Landing, Night, Clear (40)	non F-4/F-105(2.25) vs. F-105(1.90)	.012
Landing, Night, Inst (41)	non F-4/F-105(2.86) vs. F-4(2.63)	.051
Landing, Night, Inst (41)	non F-4/F-105(2.86) vs. F-105(2.43)	.004
Interdiction Mission Stress (45)	non F-4/F-105(2.93) vs. F-4(3.14)	.051
Interdiction Mission Stress (45)	non F-4/F-105(2.93) vs. F-105(3.28)	.004

[a] Compared using the Scheffe Multiple Range Test

[b] Survey scale: 1 = no stress; 2 = some stress; 3 - moderate stress; 4 - severe stress; 5 - very intense stress

[c] A comparison between fighter aircraft only

Table 3 shows the significant differences that were found among the pilots of different age groups, 21-24 (n = 74), 25-29 (n = 184), 30-34 (n = 136), 35-39 (n = 94), and 40-48 (n = 30). With the exception of five events, this table shows that in all of the comparisons the younger age groups had experienced significantly more stress than the older age groups. These data support the assumption that experience or age decreases perceived stress.

Table 3. Stress Items Showing Significant Differences Across Age at Combat Entry Groups

Variable (no.)	Groups Compared (Mean[b])	p[a]
Day Takeoff, Clear (7)	25-29(2.28) vs. 40-48(2.64)[c]	.0127
Day Takeoff, Instruments (8)	35-39(2.33) vs. 40-48(2.64)[c]	.0433
Night Takeoff, Clear (9)	21-24(2.50) vs. 25-29(2.28)	.0430
Ground Aborts (10)	21-24(1.76) vs. 40-48(2.19)[c]	.0256
4 Ship Formation (14)	21-24(1.99) vs. 40-48(1.69)	.0532
4 Ship Formation (14)	25-29(1.96) vs. 30-34(1.79)	.0382
POD Formation (16)	21-24(2.61) vs. 40-48(2.17)	.0453
POD Formation (16)	21-24(2.61) vs. 30-34(2.22)	.0103
POD Formation (16)	25-29(2.54) vs. 30-34(2.22)	.0055
Flying as Flight Cdr (18)	25-29(2.76) vs. 30-34(2.51)	.0132
Flying as Element Leader (19)	25-29(2.46) vs. 40-48(2.15)	.0467
Flying as Element Leader (19)	25-29(2.46) vs. 30-34(2.19)	.0039
Flying as Wing Man (20)	21-24(2.44) vs. 40-48(2.00)	.0148
Flying as Wing Man (20)	21-24(2.44) vs. 30-34(2.14)	.0151
Flying as Wing Man (20)	25-29(2.34) vs. 40-48(2.00)	.0366
Flying as Wing Man (20)	25-29(2.34) vs. 30-34(2.14)	.0360
Air to air: 1 Encounter (21)	21-24(3.69) vs. 40-48(3.07)	.0091
Air to air: 1 Encounter (21)	25-29(3.80) vs. 40-48(3.07)	.0008
Air to air: 1 Encounter (21)	25-29(3.80) vs. 35-39(3.38)	.0037
Air to air: 1 Encounter (21)	25-29(3.80) vs. 30-34(3.41)	.0028
Air to ground: 1 pass (28)	25-29(3.30) vs. 40-48(2.91)	.042
Air to ground: 1 pass (28)	25-29(3.30) vs. 35-39(3.01)	.023
Air to ground: 1 pass (28)	25-29(3.30) vs. 30-34(3.03)	.018
SAMs, 80mm (29)	21-24(3.97) vs. 40-48(3.51)	.017
SAMs, 80mm (29)	25-29(3.93) vs. 40-48(3.51)	.016
80mm, 57mm, and 37mm (30)	21-24(3.65) vs. 40-48(3.26)	.032
Small arms and 37mm (31)	21-24(3.19) vs. 40-48(2.66)	.006
Small arms and 37mm (31)	21-24(3.19) vs. 35-39(2.73)	.002
Small arms and 37mm (31)	21-24(3.19) vs. 30-34(2.81)	.005
Small arms and 37mm (31)	21-24(3.19) vs. 25-29(2.91)	.034
Time on target: 2 min (32)	21-24(3.34) vs. 40-48(2.88)	.031
Time on target: 4 min (33)	21-24(3.57) vs. 40-48(3.03)	.0134
Time on target: 4 min (33)	25-29(3.54) vs. 35-39(3.28)	.0480
Time on target: 6 min (34)	21-24(3.93) vs. 40-48(3.33)	.0108
Time on target: 6 min (34)	21-24(3.93) vs. 30-34(3.60)	.0444
Time on target: 6 min (34)	25-29(3.90) vs. 40-48(3.33)	.007
Time on target: 6 min (34)	25-29(3.90) vs. 30-34(3.60)	.0181
Landing, night, inst (41)	21-24(2.86) vs. 30-34(2.52)	.0205

Variable (no.)	Groups Compared (Mean[b])	p[a]
Landing, night, inst (41)	21-24(2.86) vs. 35-39(2.52)	.0311
Ground Support Mission (46)	25-29(2.41) vs. 40-48(2.79)[c]	.0522
Ground Support Mission (46)	35-39(2.37) vs. 40-48(2.79)[c]	.0448
Air Cover Mission (47)	21-24(2.78) vs. 40-48(1.96)	.0014
Air Cover Mission (47)	21-24(2.78) vs. 30-34(2.41)	.0325
Air Cover Mission (47)	25-29(2.51) vs. 40-48(1.96)	.0189

[a] Compared using the Scheffe Multiple Range Test.

[b] Survey scale: 1 - no stress; 2 - some stress; 3 - moderate stress; 4 - severe stress; 5 - very intense stress.

[c] These comparisons show the younger age group with less stress than the older age group.

Conclusions

The pilots surveyed appear to agree in their perceptions of which mission events are most stress-provoking. However, the F-4 and F-105 pilots seem to perceive the very dangerous combat events; i.e., SAMs, anti-aircraft artillery, and Mig's, as more stressful than the non F-4/F-105 pilots, even though it is known that these F-4 and F-105 pilots encountered these events more often than the non F-4/F-105 pilots. These results may support Shaffer's research since the length of exposure, although incremental, to the greatest stress-provokers is longest for the F-4 and F-105 pilots. However, the incremental effect appears to be a somewhat different problem than a single exposure to combat.

Very few differences were found in comparing F-4 to F-105 pilots. This implies that the extra pilot, extra engine, and better performance characteristics of the F-4 versus the F-105 have little effect on combat stress perception. This contradicted the expected results, the extra pilot (the F-4 has a two seat, front-back arrangement) particularly was expected to help decrease combat stress.

This study also indicates that the older person perceives less stress in the air combat environment. This conclusion does not particularly contradict Shaffer's work because the older pilot did not necessary have more hours of combat experience. Future research using these data will assess the relationships between stress events and sortie estimates for various pilot group, aircraft types, and mission profiles.

References

Austin, F.H., Jr. A review of stress and fatigue monitoring of naval aviators during aircraft career combat operations: Blood and urine biochemical studies. In Bourne, P.G. (Ed.) The psychology and physiology of stress with reference to special studies of the Vietnam war. New York: Academic Press, 1969.

Kantor, J.E., Klinestiver, L., & McFarlane, T.A. Methodology to assess psychological stress and its impact in the air combat environment (AFHRL-TR-78-3). Brooks AFB TX: Air Force Human Resources Laboratory, Personnel Research Division, 1978.

Shaffer, L.F. Fear in combat and its control. Washington, D.C.: The Working Group on Human Behavior Under Conditions of Military Service, Department of Defense, Research and Development Board, 1951.

Interpersonal Stress as a Moderator of the Relationship Between Intelligence and Academic Performance

Earl H. Potter
United States Coast Guard Academy
New London, Connecticut 06320

Abstract

This study investigates the role which interpersonal stress plays in the relationship between academic aptitude and academic performance. The subjects are 107 cadets at the United States Coast Guard Academy. The results suggest that the academic performance of junior cadets (freshmen and sophomores) is related to their academic aptitude (Scholastic Aptitude Test-math score) when interpersonal stress between cadets is low. For senior cadets (juniors and seniors) the relationship between academic performance and aptitude does not vary with stress between cadets. For senior cadets the relationship of aptitude to performance is greater when there is stress between the cadet and his company officer and/or instructor. These findings are related to current theories of stress and anxiety, as well as to previous studies of the effects of stress on intellectual performance. Treatments which may reduce the detrimental effects of stress on junior cadets are suggested.

This study has its roots in an organizational conondrum. Despite the commonly held assumption that intelligent persons will perform most jobs better than less intelligent persons, research studies consistently report that intelligence does not make a significant contribution to job performance (e.g. Mann, 1958; Stogdill, 1974). These findings are puzzling, first because one has to assume some institutional wisdom which has led to the extensive use of intelligence as a selection criterion. Second, one can identify elements of many jobs which require the individual to abstract, conceptualize, and make complex judgements. All of these are functions encompassed by the usual definitions of intelligence (Butcher, 1968). By all rights, then, intelligence ought to correlate with performance.

A possible explanation for the observed low correlations of job performance and intelligence is that intelligence contributes to performance under some conditions and not under others (Fiedler & Leister, 1977). One factor which the psychological research strongly suggests as a possible moderator of the relationship between intelligence and performance is stress. Lazarus (1966) reports that anxiety and stress narrow the individual's focus and inhibit creative thought. Amabile (1979) likewise finds that individuals apprehensive about evaluations of their work perform less well on creative tasks. In a recent series of studies of military leadership (Fiedler, Potter, Zais, and Knowlton, 1979), stress between the leader and his or her superior was found to moderate the relationship between leader intelligence and leader performance. Specifically, when stress with the leader's boss was high the correlation between intelligence and performance was low or, in some cases, negative. When stress with the boss was low the correlation of leader intelligence and leader performance was high.

PRECEDING PAGE BLANK-NOT FILMED

Another factor which may moderate the relationship to performance is job experience. If stress results from a difficult demand being placed on an individual (McGrath, 1976), experience may, by providing an array of possible coping strategies, compensate for the effects of stress. Such seems to be the case in at least one study (Potter, 1978), in which the effects of stress are stronger when experience is low than when it is high.

These findings suggest that in other situations where intelligence ought to contribute to performance a preoccupation with stressful interpersonal relationships might also block the contribution of intelligence to performance. Recent research on the effects of anxiety on academic performance (Tobins, 1979) was focuses primarily on instructional methods or personality traits, such as test anxiety. While there is evidence that the student's environment can influence academic performance (Blinberg & Hample, 1979), little attention has been paid to sources of anxiety or stress outside of the classroom. This study examines the impact of interpersonal stess on the contribution of intellectual ability to academic performance for students who are low in experience (freshmen and sophomores and students who are high in experience (juniors and seniors). The following hypotheses will be tested.

Hypothesis I

For freshmen and sophomores the relationship between intelligence and academic performance will be lower under conditions of high interpersonal stress than under conditions of low interpersonal stress.

Hypothesis II

For juniors and seniors there will be a significant positive correlation between intelligence and performance under conditions of high interpersonal stress as well as under conditions of low interpersonal stress.

Method

Subjects. Subjects were 107 cadets enrolled in a required psychology course at the United States Coast Guard Academy. Of the cadets participating in the study, 35 were seniors, 23 juniors, 44 sophomores, and 5 freshmen. The great majority of subjects were white (105) males (99).

Measurements

Intellectual Ability. The curriculum at the U.S. Coast Guard Academy is predominantly an engineering curriculum. Hence, the mathematical scores of the Scholastic Aptitude Test (SAT-M) was chosen as the most relevant indicator of intellectual ability. For this sample scores range from 500 to 780. The mean SAT-M score was 655.2 with a standard deviation of 62.66.

Academic Performance. Academic performance was indicated by the cadets cumulative grade point average (GPA). This score is the average of the unweighted products of the course grades multiplied by the credit hour value of each course. GPA's for this sample ranged from 1.42 to 3.96 with a mean of 2.83 and a standard deviation of .55.

Interpersonal Stress. Interpersonal stress was determined for five relationsnips considered to be important to most cadets. These were the relationships with parents, peers, (i.e. classmates and to some extent cadets junior in class), cadets senior in class, the company officer, and the several academic instructors. The four item scale used to assess stress in these relationships was based on the factors determined by Potter (1978) to be important consituents of stress in the relationship between a leader and the leader's boss. In that relationship stress is most dis-

ruptive when that stress results from the boss making significant demands but withholding guidance and support from the subordinate. Under these conditions subordinates feel uncertain about what they should do, on the one hand, and under pressure to do something, on the other. It was reasoned that this same situation would be stressful if it existed in other significant relationships. Therefore, cadets were asked to consider one of the significant relationships and asked to indicate on a scale of 1 to 7 how much stress or tension they felt as a result of certain behaviors. For example: "He rates me low in aptitude but he doesn't tell me what I should do in order to improve my grade."

The four stress items resulted in a one factor scale with a possible range of 4 to 28. Mean scores varied from a low of 7.65 (range = 4-21, SD = 3.45) for stress with parents to 13.04 (range = 4-24, SD = 5.4) for stress with senior cadets. Means of the other stress measures were 11.33 (range = 4-26, SD = 5.16) for stress with peers, 11.85 (range = 4-26, SD = 6.13) for stress with instructors, and 12.76 (range = 4-28, SD = 6.17) for stress with company officer.

Results

An examination of the intercorrelations among all variables shows no significant relationships except for a significant correlation between SAT-M and GPA for all subjects ($r = .51$, $N = 106$, $p < .001$) and significant moderate correlations among the stress measures.

Hypothesis I

Hypothesis I was supported only with respect to stress with peers and senior cadets. Freshmen and sophomores were split approximately in thirds to represent high, low and moderate stress in each relationship. For subjects who reported high or low stress with parents, company officers, or instructors there was no significant difference in the SAT-M & GPA relationship. For subjects who reported high or low or moderate stress with peers or senior cadets the difference in the SAT-M & GPA realtionship were significant and striking. For cadets who reported high stress with peers the relationship of intelligence to performance was not significant ($r=.08$, $N=15$). For cadets who reported moderate stress this relationship was significant ($r=45$, $N=18$, $p < .05$). For cadets who reported low stress with peers the relationship was also significant ($r=.76$, $N=14$, $p < .001$). The difference between the intelligence/performance correlation for cadets with high and low stress was significant using a Fisher's Z test. ($p < .05$).

For cadets reporting high, moderate, or low stress with senior cadets the picture was the same: high stress ($r=.04$, $N=15$, NS), moderate stress ($r=60$, $N=19$, $p < .01$), high stress ($r=76$, $N=13$, $p < .001$). A Fisher's Z test shows a significant difference between the intelligence/performance relationships for high and low stress with senior cadets ($p < .05$).

Clearly, not all relationships are equally salient to junior cadets. For those cadets who are relatively new to the Academy system their personal relationships in the barracks are most important. If freshmen and sophomore cadets are preoccupied with trying to meet the demands of senior cadets they appear unable to apply their intellectual ability to their academic tasks.

Hypothesis II

Hypothesis II received mixed support but the pattern of results offers some suprising suggestions. For subjects who reported high or low stress with senior cadets, company officers or instructors, the relationship of intelligence to performance was significant and positive under conditions of both high and low stress. However, contrary to the pattern observed with junior cadets, the relationship was stronger for cadets who reported high stress with either the company officer ($r=.64$, $N=19$, p .01) or the instructor ($r=.78$, $N=.18$, p .001). This finding may be comparable to the findings of Zajonc (1976) who reports that when a task is practiced stress results in increased performance. It may be that stress for senior cadets reflects motivation and that their greater experience has given them answers to their problem, while junior cadets have only problems. Support for this explanation is provided by Potter & Fiedler (in press) who found that under conditions of high stress with the boss experience contributed significantly to performance while under conditions of low stress with the boss, this was not the case.

For subjects who reported high or low stress with parents or peers the intelligence performance relationship was significant under conditions of low stress ($r=.61$, $N=20$, p .01 and $r=.55$, $N=18$, p .02) and approached significance for subjects who reported high stress ($r=.42$, $N=19$, p .06 and $r=.40$, $N=21$, p =.06). For subjects reporting high or low stress with senior cadets the results were similar but reached significance ($r=.47$, $N=18$, p .05 and $r=.56$, $N=19$, p .01). It would appear then that experience has taught senior cadets how to deal with life in the barracks. It would also appear that the contribution of SAT-M to GPA is less significant than in the first two years, however, these differences are not significant.

Discussion

This study shows that stress in significant relationships moderates the contribution of intelligence to the academic performance of college students. For students who lack experience in the institution and hence specific skills to cope with the stressor, a preoccupation with interpersonal demands results in a decrease in the contribution of intelligence to performance. For students who through tenure have gained coping skills stress does not decrease the contribution of intelligence to performance. On the contrary, demands made by instructors and company officers which are perceived as stressful may even result in an increased contribution of intelligence to performance. Presumably these students motivated by increased pressure apply their intelligence to solving the problems with which they are confronted. If experience guides the process then fewer mistakes and misapplications may be made resulting in a greater contribution being made by the students intelligence.

For the academic administrator concerned with optimizing the performance of students this study offers some clues. It is uncertainty in significant relationships that causes the most bothersome stress. Surprisingly these relationships are not only student-teacher relationships. Sarason (1979) suggests that self-preoccupation which results from stress causes a person to misdirect efforts away from the primary task.

Quick (1979) suggests that setting clear goals reduces stress due to uncertainty. Taken together these findings point to the need for increasing clarity and setting clear goals and a means to assist less experienced students in reducing misdirected efforts. If sources of stress are reduced for these students, they should be better able to apply their abilities to the task. The contribution of their intelligence to their performance should go up and the resulting experience of success should reinforce the task efforts that have been made.

References

Amabile, T.M. Effects of external evaluation on artistic creativity. Journal of Personality and Social Psychologist, 1979, 37, 221-233.

Bliming, G.S., & Hample, D.H. Structuring the Peer Environment in Residence Halls to Increase Academic Performance in Average-Ability Students. Journal of College Student Personnel, 1979, 20, 310-316.

Butcher, H.J. Human Intelligence: Its nature and assessment. New York: Harper & Row, 1968.

Fiedler, F.E., & Leister, A. Intelligence and group performance: A multiple screen model. Organizational Behavior and Human Performance, 1977, 20, 1-14.

Fiedler, F.E., Potter, E.H., III., Zais, M.M., & Knowlton, W.A., Organizational Stress and the Use and Misuse of Managerial Intelligence and Experience. Journal of Applied Psychology, 1979, 64, 635-647.

Lazarus, R.S. Psychological stress and the coping process. New York: McGraw-Hill, 1966.

Mann, R.D. A review of the relationships between personality and performance in small groups. Psychological Bulletin, 1959, 56, 241-270.

McGrath, J.E., Stress and behavior in organizations. N. Dunnette, M.D., (editor). Handbook of Industrial and Organizational Psychology; Chicago; Rand-McNally, College Pub. Co., 1976.

Potter, E.H., III. "The contribution of intelligence and experience to the performance of staff personnel." Unpublished doctoral dissertation, University of Washington, 1978.

Potter, E.H., III & Fiedler, F.E. "Stress and the utilization of staff member intelligence and experience." Academy of Management Journal, in press.

Quick, J.C. "Dyadic goal setting and role strss: A field study." Academy of Management Journal, 1979, 22, 241-252.

Sarason, I.G. Life stress, self-preoccupation, and social supports. Technical papers. Office of Naval Research, 1979.

Stogdill, R.M. Manual for the Leader Behavior Description Questionnaire--Form XII: An experimental revision. Columbus: Ohio State University, 1963.

Tobias, S. Anxiety Research in Educational Psychology. Journal of Educational Psychology, 1979, 71, 573-582.

Zajonc, R.R. Social facilitation. Science, 1965, 149, 269-274.

The Differential Effectiveness of Specific Treatments on the Worry-Emotionality Components of Test Anxiety

Kermit E. Parker, Jr.
Wilford Hall USAF Medical Center

Abstract

The purpose of this study was to determine the effects of a cognitive-attentional treatment and a variation of systematic desensitization on the worry-emotionality components of test anxiety. Forty-six high test anxious undergraduate college students were divided into three levels of study skills ability. Subjects within each study skills level were then assigned to one of the following treatment conditions: a cognitive-attentional condition; a coping systematic desensitization condition; and a no treatment control condition. Results indicated that each treatment condition showed significantly greater reductions in levels of test anxiety than the no treatment control condition. However, there were no performance increases evidenced by the treatment or control conditions. It was concluded that either treatment would effectively reduce high levels of test anxiety but that these treatments were ineffective in increasing academic performance. The relationship between test anxiety and performance was questioned and alternative viewpoints discussed.

Research on test anxiety has centered around the description and identification of conditions which separate high and low test anxious individuals. One major effort which further characterized test anxious persons was accomplished by Liebert and Morris (1967; Morris & Liebert 1969, 1970) in a series of factor analytic studies. From this research these authors proposed that test anxiety contained two major components. One of these components, termed worry, was considered to consist of anxiety mediated task irrelevant responses manifested in the form of self-centered derogatory type statements, such as "I am stupid" or "What if I fail?". These type responses exemplified the distracting and attentionally demanding characteristic of worry cognitions prevalent in high test anxious persons and were posited to be the contributing factor in decreased performance (Holroyd, 1976; Wine, 1971). The second component, termed emotionality, was proposed to consist primarily of heightened autonomic arousal to the examination situation and was considered to have little influence on performance except when worry was very low.

This theoretical formulation had a significant input into the applied or treatment aspects of test anxiety. The treatment of choice for test anxiety has been primarily systematic desensitization (Allen, 1972). However, systematic desensitization as a treatment procedure functions to reduce autonomic arousal or emotionality, which was considered to have the least detrimental effect upon performance. As such, Wine (1971) concluded that treatments of test anxiety should focus upon reducing the worry aspects of test anxiety if performance increases are to be expected. The few studies which have examined treatment procedures with this focus (Holroyd, 1976; May, 1976; Meichenbaum, 1972; Wine, 1971) have met with equivocal results. One major flaw of these and other treatment studies was

PRECEDING PAGE BLANK-NOT FILMED

that changes in worry and emotionality were not measured. It therefore remained unclear whether treatments designed to reduce a particular component of test anxiety were able to demonstrate performance increases as a result of reductions in a specific component of test anxiety.

In an attempt to clarify some of the theoretical and applied issues of test anxiety research, the present study attempted to examine the differential effectiveness of two treatment procedures specifically targeted to reduce levels of each component of test anxiety. In addition, this research attempted to determine what effect, if any, these reductions would have on increased academic performance.

Method

Subjects

Subjects for this experiment were undergraduate college students attending two private Southern universities. The Test Attitude Inventory (TAI) (Spielberger, Gonzalez, Taylor, Algaze & Anton, 1978) was administered to introductory psychology classes at both schools. Students who scored above the 63rd percentile on the total score of the TAI, excluding freshmen, were invited to participate. A total of 60 individuals met these criteria and 44 students volunteered for the study. Five subjects, who learned of the experiment through other sources volunteered to participate and were accepted for the study by meeting the same criteria. Three subjects dropped out of the study prior to the first treatment session leaving a total of 46 subjects who completed the study.

Materials

Materials required for this experiment consisted of the following:

1. The Test Attitude Inventory (Spielberger et al., 1978) which consisted of total, worry and emotionality scales.
2. The EDL Reading Test which was used as an immediate performance measure.
3. The Brown-Holtzman Survey of Study Habits and Attitudes (SSHA) was used to determine each subject's level of study ability.

Procedure

Study skills and test anxiety are negatively correlated. Thus it was expected that most subjects in this study would have poor study ability. Seventy-two percent of the subjects participating in this study scored below the 50% level on the SSHA indicating this was indeed the case. However, due to experimental design constraints subjects were divided into three equal groups based on total scores on the SSHA. Scores for the three groups ranged from 0-15% for the low study skills group, 20-30% for the moderate study skills group and 35-95% for the high study skills group. Subjects were divided in this manner to provide a statistical control for the different levels of study skills among the subjects although the distribution of study ability. Subjects within each study skills level were then assigned randomly to one of three treatment conditions which consisted of the following: a cognitive-attentional condition, a coping systematic desensitization condition, or a no treatment control condition.

Treatments were conducted in group sessions consisting of from four to eight subjects. Each group received two treatment sessions per week for a total of eight sessions. Each session was approximately 50 minutes in length.

The cognitive-attentional condition was based on Wine's (1971) model and followed the procedures utilized by Holroyd (1976). The purpose of this treatment procedure was to reduce the maladaptive cognitive responses and task-irrelevant ruminations by making the test anxious person aware of the negative self-statements and distracting thoughts made prior to and during an examination. Additionally, subjects were trained to emit incompatible self statements designed to facilitate attention to the task accompanied by suggestions for alternate methods of interpreting and labeling emotional arousal.

The coping systematic desensitization procedure was utilized by Meichenbaum (1972) as a component of a cognitive modification procedure. The procedure followed a standard systematic desensitization format of basic relaxation training, group hierarchy construction and imagery training. For the initial hierarchy items the standard mastery technique was used. During the presentation of the remaining hierarchy items the coping procedure was used in one of two ways. On some items the therapist directly suggested the subject would become anxious during the presentation of the item. When an item was presented in this manner the therapist included in the image appropriate relaxation coping strategies. For the last scenes of the hierarchy the therapist varied between the mastery and coping imagery. A standard test anxiety hierarchy (Deffenbacher, 1974) was provided as a guide to each systematic desensitization treatment group but the hierarchy was modified according to the needs of each group. The procedure followed a group desensitization format with the group progressing through the hierarchy at the rate of the slowest member. Group members completed all items in each hierarchy.

Subjects in the no treatment control condition came in only for the three assessment sessions and had no other contact with the experimenter during this time.

Post-treatment data for the TAI and EDL test were collected during the last treatment session (session eight). In keeping with the view that some passage of time was required to integrate new skills and behavior, one week prior to final examinations (five weeks from the last treatment session) each subject was asked to return to complete the post-treatment battery once again. End of semester GPA was also collected for each subject and compared with the cumulative GPA at the beginning of the semester.

Results

The data collected for each dependent measure at three different points in time (Pre-treatment, post-treatment, followup) were analysed using a 3 X 3 X 3 (study skills X treatment X time) repeated measures ANOVA.

The primary analysis for this study was focused upon the total, worry and emotionality scales of the TAI. Each scale was analysed separately. Results for the total test anxiety scale, using normalized t scores revealed a significant decrease from pre-post assessment ($p < .01$), from pre-assessment to followup ($p < .01$), but no significant difference in scores from post assessment to followup for the cognitive treatment group. The coping systematic desensitization group showed a significant decrease from pre-post treatment ($p < .01$), from pre-treatment to followup ($p < .01$) but no significant difference from post treatment to followup. Subjects in the control group showed no significant decrease in total test anxiety scores across the three

assessment periods.

There were no significant differences between the three treatment groups on TAI total scores at the pre-treatment assessment period. At the post treatment period the cognitive and coping systematic desensitization treatment groups differed significantly from the control group ($p < .01$ for both) but were not significantly different from each other. This same pattern of results was present for the followup assessment period as well.

Results for the TAI worry scale revealed decreases in scores across the three assessment periods similar to those of the total scale. Both the cognitive and coping systematic desensitization treatment groups showed significant decreases from pre-post treatment ($p < .01$, $p < .05$), from pre-treatment to followup ($p < .01$ for both) but no difference from post treatment to followup. Subjects in the control group showed no significant decrease in worry scores across the assessment periods. There were no significant differences in worry scores between the treatment groups for the pre and post treatment periods. At the followup period the cognitive and coping systematic desensitization groups differed significantly from the control group ($p < .05$ for both) but there were no significant differences between these two groups.

For the TAI emotionality scale the results across assessment periods were consistent with those found for the total and worry scales. Between group comparisons revealed significant differences between both the cognitive and coping systematic desensitization groups and the control group for the post treatment and followup assessment periods. Again, there were no significant differences between the cognitive and coping systematic desensitization groups for any of the assessment periods.

Performance measures used in this study consisted of the EDL reading test (forms C and D) and GPA. On the EDL test all three groups showed an increase in performance from pre-treatment to followup ($p < .05$), from post treatment to follow ($p < .01$) but no increase from pre to post assessment. All other main effects and interactions for this measure were not significant. Analysis of GPA data, using pre-semester cumulative GPA and end of semester GPA as the pre-post comparison revealed no significant main or interaction effects.

Discussion

One purpose of this study was to determine the effects of specific test anxiety treatments on worry and emotionality. The results indicated no differential treatment effectiveness. Each treatment procedure produced similar reductions in total, worry and emotionality scores as measured by the TAI. The cognitive-attentional treatment group did show significant decreases in TAI total, worry and emotionality scores across the three assessment periods as expected. This decrease was significantly different from the control group but was no greater than the decrease displayed by the coping systematic desensitization group. The coping systematic desensitization group did show significant decreases in emotionality scores, however, contrary to expectations this group showed significant reductions for both TAI total and worry scores as well.

The evidence from this study revealed that the experimental groups were in fact different from the control group indicating clearly that some form of treatment was more effective than no treatment in reducing levels of test

anxiety as measured by self-report indices. However, neither treatment strategy was demonstrated to be more effective than the other. Analysis of experimental variables indicated that neither experimental confounding nor general treatment effectiveness of the treatment approaches was able to account for the lack of differentiation in treatment effectiveness. Therefore, it seemed that the original theoretical views should be reexamined.

The measurement of specific changes in worry and emotionality provided empirical evidence that all aspects of test anxiety can be reduced by a given treatment procedure and not singular reductions in specific factors as suggested by theory. It was concluded from these results that it may not be necessary to conceptualize test anxiety as multi-faceted. Much of the theoretical literature about test anxiety was based upon the dual aspect proposed by Liebert and Morris (1967). It seemed that more importance has been attached to viewing test anxiety in terms of worry and emotionality than was warranted. These factors of test anxiety were presented as separate entities having unique qualities in and of themselves which required specialized treatment procedures to reduce their detrimental effects. Based on the results of this study, it appeared that no functional purpose was served by conceptualizing test anxiety in this manner. When treatment effectiveness was evaluated, what the focus of the treatment was did not seem to matter because all aspects of test anxiety were reduced.

The results for the performance measures were contrary to expectations, although typical of test anxiety research. There were no performance increases which could be attributable to treatment effectiveness on either the EDL test or GPA. The results of this and most studies of test anxiety reduction have not shown performance increases as a result of reducing high levels of test anxiety. Spielberger et al. (1978) indicated that "there is little evidence that behavioral treatments alone can facilitate academic achievement" (p. 20). In light of these findings it appeared that the effects of test anxiety upon performance must be reevaluated. Motivation and study skills are two important factors in the academic environment which probably inhibit performance to a greater extent than does test anxiety. Wolpe, Brady, Serber, Agras and Lieberman (1973) indicated that an individual must already possess the necessary skills for coping with an anxiety related problem before an anxiety reduction method would be effective. Deficient study skills and poor motivation would appear to be significant problems among test anxious students which must be remedied before academic performance gains can be expected. In addition, it appears that an individual must achieve or possess a certain level of study skills and motivation before a treatment for test anxiety is appropriate.

The evidence from this study permitted the following conclusions to be drawn: (a) That a behaviorally oriented treatment approach, either systematic desensitization or cognitive modification, will effectively reduce high levels of test anxiety associated with examination stress; (b) a test anxiety reduction strategy does not appear to be effective in increasing academic performance; and (c) that intervention in other areas such as motivation and study skills will more likely lead to academic performance increases.

References

Allen, G. J. The behavioral treatment of test anxiety: Recent research and future trends. Behavior Therapy, 1972, 3, 253-262.

Deffenbacher, J. L. Hierarchies for the desensitization of test and speech anxiety. Journal of College Student Personnel, 1974, 15, 452-454.

Holroyd, K. A. Cognition and desensitization in the group treatment of test anxiety. Journal of Consulting and Clinical Psychology, 1976, 44, 991-1001.

Liebert, R.M., & Morris, L.W. Cognitive and emotional components of test anxiety: A distinction and some initial data. Psychological Reports, 1967, 20, 975-978.

May, R. L. The treatment of test anxiety by cognitive modification: An examination of treatment components. Dissertation Abstracts International, 1976, 37, 468-469.

Meichenbaum, D. H. Cognitive modification of test anxious college students. Journal of Consulting and Clinical Psychology, 1969, 33, 240-244.

Morris, L. W. & Liebert, R. M. Relationship of cognitive and emotional components of test anxiety to physiological arousal and academic performance. Journal of Consulting and Clinical Psychology, 1970, 35, 332-337.

Spielberger, C. D., Gonzalez, H. P., Taylor, C. J., Algaze, B., & Anton, W. D. Examination stress and test anxiety. In C. D. Spielberger & I. G. Sarason (Eds.), Stress and anxiety, (Vol. 5). New York: Hemisphere/Wiley, 1978.

Wine, J. Test anxiety and direction of attention. Psychological Bulletin, 1971, 76, 92-104.

Wolpe, J., Brady, J. P., Serber, J., Agras, W. S., & Lieberman, R. P. The current status of systematic desensitization. American Journal of Psychiatry, 1973, 130, 916-965.

Test Anxiety and Physiological Response to Evaluation Stress

Scott W. Carrothers and Richard L. Hughes
United States Air Force Academy, Colorado

Abstract

One assumption underlying desensitization approaches to treating test anxiety is that test-anxious persons are, in fact, muscularly tense during stressful evaluative experiences. The present study investigated frontalis muscle tension and tension and performance self-reports of high and low test-anxious Ss in stress and nonstress situations. There were no significant differences between frontalis tension levels between high and low test-anxious subjects in either stress condition. This supports the view that test anxiety is due to cognitive factors and not physiological overarousal. Further support of this view was demonstrated by the differences between the anxiety groups in the tension and performance self-reports. The high test-anxious Ss reported significantly higher subjective tension levels and lower estimations of personal performance on the tracking task.

For well over a decade the dominant mode of treatment of test anxiety has been systematic desensitization or some variant of it. In all of these treatments relaxation training has played a major part, whether that be through progressive relaxation, autogenic training, or biofeedback. Many researchers (e.g., Liebert & Morris, 1967) conceptualize two factors of test anxiety - an emotionality or physiological factor and a worry or cognitive factor - but most treatments until recently have emphasized the former. Most treatments have attempted to lower physiological arousal in test anxious subjects rather than to modify their dysfunctional cognitions about testing. However, the effectiveness of relaxation treatments in reducing test anxiety often was not matched with comparable improvement in posttreatment test performance (Spielberger, Anton, & Bedell, 1976). Cognitive modification procedures, on the other hand, have proven effective in reducing test anxiety as well as in improving posttreatment academic performance (Holroyd, 1976).

The relative success of cognitive modification procedures in improving performance suggests that dysfunctional cognitive and attentional processes may be more central in test anxiety than is emotionality or overarousal. Wine (1971) suggested this and Meichenbaum's (1972) work supports it. Furthermore, there is increasing evidence that physiological overarousal in testing situations does not, after all, characterize test anxious subjects. Recent studies have failed to document any actual differences in the physiological arousal of low and high test anxious subjects despite clear differences in their self-reports of arousal.

For example, Hughes (1979) measured the frontalis muscle tension of high and low test anxious subjects during the imaginal visualization of stressful testing situations and there were no differences between the groups. In addition, the actual electromyographic (EMG) tension levels in both groups suggested a relatively relaxed musculature rather than overarousal. Both of these findings are inconsistent with the view that test anxiety is caused by physiological overarousal. In another study, Holroyd, Westbrook, Wolf, & Badhorn (1978) studied physiological response during

testing situations and found no differences between high and low test anxious subjects on many autonomic indices including electrodermal activity and heart rate.

The present study was designed to further explore the relationship between test anxiety and physiological response. The study differed from Holroyd et al by examining frontalis (EMG) levels instead of autonomic indices, and it differed from the Hughes study by examining frontalis tension in actual testing situations which varied in stressfulness as a function of the type of instructions presented to the subjects.

Procedure

Approximately 500 students in an introductory psychology course completed the Test Anxiety Inventory (Spielberger). Groups of extremely high and extremely low test anxious subjects were selected for the experiment on the basis of their scores on this instrument.

A critical tracking task (Jex, McDonnell & Phatak, 1966) was the test situation in this experiment. It was introduced to high and low test anxious subjects in either a relaxed and nonevaluative manner or a formal and highly evaluative manner. Instructional conditions have previously been shown to be an important variable affecting the performance of test anxious subjects. Frontalis EMG was recorded during a five-minute baseline period and also during the five minute task performance itself. In addition to EMG levels, other dependent variables included self-reports of subjective tension, tracking scores, and estimations of task performance.

Results

There were no significant differences in the mean frontalis tension levels during performance of the task between the low ($\bar{X}$ = 3.96 microvolts/minute) and high ($\bar{X}$ = 4.85 microvolts/minute) test anxiety groups. It should be noted, however, that this difference was in the expected direction and only barely missed conventional levels of significance ($p = .07$). A much more significant effect ($p < .001$) was noted in the subjects descriptions of their felt tension. The high test anxious subjects described themselves as much more tense than did the low anxious subjects. The latter difference in self-labeling between the two groups was also noted in their performance self-reports. When subjects were asked to compare their performance with that of others, without having any acutal data on norms for the task, the test anxious subjects significantly ($p < .05$) devalued their performance relative to others. This was in spite of the fact that their actual performance levels were not different. There were no significant effects due to the stressfulness of instructions. These results all lend support to cognitive formulations of test anxiety.

References

Holroyd, K. A. Cognition and desensitization in the group treatment of test anxiety. Journal of Consulting and Clinical Psychology, 1976, 44, 991-1001.

Holroyd, K. A., Westbrook, T., Wolf, M., & Badhorn, E. Performance, cognition, and physiological responding in test anxiety. Journal of Abnormal Psychology, 1978, 87, 442-451.

Hughes, R. L. Test anxiety, electromyographic response to stress and reflective gaze preference. FJSRL Technical Report 79-0009, Project 2303. Frank J. Seiler Research Laboratory, Air Force Systems Command, 1979.

Jex, H. R., McDonnell, J. D., & Phatak, A. V. A "critical" tracking task for man-machine research related to the operator's effective delay time. NASA CR-616, Oct. 1966.

Liebert, R. M., & Morris, L. W. Cognitive and emotional components of test anxiety: a distinction and some initial data. Psychological Reports, 1967, 20, 975-978.

Meichenbaum, D. H. Cognitive modification of test-anxious college students. Journal of Consulting and Clinical Psychology, 1972, 39, 370-380.

Spielberger, C. D., Anton, W. D., & Bedell, J. The nature and treatment of test anxiety. In M. Zuckerman & C. D. Spielberger (Eds.) Emotion and Anxiety: New Concepts, Methods and Applications. New York: Wiley,

Wine, J. Test anxiety and direction of attention. Psychological Bulletin, 1971, 76, 92-104.

Effect of Alternate Instrument Displays on Novice Pilots

Ross A. Cannon and Jock C. H. Schwank
Department of Behavioral Sciences and Leadership
USAF Academy, Co 80840

Abstract

This experiment investigated the effectiveness of using peripheral light displays to indicate heading during flight in a GAT-1 simulator. Twenty male and 20 female USAFA cadets with no previous flight experience flew the simulator for one hour. There were only minor performance differences between males and females. In general, both groups were able to fly the GAT-1 at least as well when using peripheral displays as when the normal round dial displays were used. The results of this experiment are consistent with previous experiments using more flight-experienced subjects and different peripheral displays.

This report details the latest in a continuing series of experiments designed to test recent theoretical developments concerning information processing characteristics of human vision and attention. Specifically, this experiment investigates the effectiveness of presenting aircraft control information in the extreme periphery of the visual field rather than using normally placed foveal instrument displays.

Previous studies in this series have shown that using peripheral lights yield no significant decrement in flight performance when compared to using normal foveal instrument displays (Schwank, et al, 1978; Godfrey, et al, 1978; Bermudez, et al, 1979; Nataupsky, et al, 1979). Also, it has been established that steady peripheral displays yield better performance than peripherally placed vertical strobing lights (Schwank, et al, 1978; Godfrey, et al, 1978; Bermudez, et al, 1979). Previous studies have also shown that subjects crossed through the desired heading more frequently (overcorrections) using peripherally located displays than when using normal instrument displays (Schwank, et al, 1978; Godfrey et al, 1978; Bermudez, et al, 1979).

The purpose of the current experiment was to investigate the effects of sex differences in the use of peripheral displays. With the advent of more women entering pilot training, it was felt important to determine whether women could use peripheral signals as well as men. Also, this experiment investigated novice pilots' ability to use alternate instrument displays. T-41 qualified pilots have been used in the three previous simulator studies in this series (Schwank, et al, 1978; Godfrey, et al 1978; Bermudez, et al, 1979). Novice pilots were used to eliminate the possibility of proactive inhibition resulting from prior flying experience with normal instrument displays. Finally, this experiment used a horizontal strobing light versus the vertical strobing light used in previous studies. It is possible that a horizontal strobe could be a more compatible stimulus for the required responses (i.e., turning left or right) than the vertical strobe.

PRECEDING PAGE BLANK-NOT FILMED

Method

Subjects

Subjects were 40 volunteers from the United States Air Force Academy who had never piloted an aircraft. All subjects were between the ages of 18 and 24. Of the 40 subjects, half were male and half female. Male students were randomly divided into one of two groups, one group of 10 using a magnetic heading indicator and the other group of 10 using peripheral lights to indicate heading. The women were also randomly assigned to form two groups of 10 with each group using either a magnetic heading indicator or peripheral lights.

Design

A diagram of the experimental design is presented below.

Table 1
Experimental Design

Sex	Indicator	
	Magnetic	Peripheral
Male	S_1	S_{21}
	.	.
	.	.
	.	.
	S_{10}	S_{30}
Female	S_{11}	S_{31}
	.	.
	.	.
	.	.
	S_{20}	S_{40}

Both factors, heading indicator and sex, provided between subjects experimental groups. The resulting design was a 2 X 2 analysis of variance with no repeated measures.

Apparatus

Training and experimental sessions were conducted using a Singer-Link GAT-1 Simulator, Model 633000. This aircraft simulator provides the same primary instrument display as the T-41 aircraft (heading indicator, airspeed indicator, vertical velocity indicator, altitude indicator, attitude indicator, tachometer, and turn and slip indicator). Cockpit design and simulated flight characteristics are also similar to the T-41.

For peripheral heading cues, the pilots wore a faceplate which held two banks of light emitting diodes (LEDs). Each bank consisted of five LEDs, green in color and wired in series. Each LED was 2.4mm in diameter. The five LEDs were placed 6 mm apart in a horizontal line. This visual display yielded a horizontal strobe of light which was rate-yoked to heading deviation (i.e., as the simulator moved farther from the desired heading of 270^{o}, the lights strobed at a higher frequency). Onset of the light occurred only when the pilot exceeded the desired heading by ± 1 degree. The strobe signal was positioned such that the center LED of the five-light sequence was on the horizontal meridian (0^{o}) of the nasal retina at 55 degrees of eccentricity from the fovea when the pilot fixated a point along his normal viewing axis at a distance of three meters. Total visual angle

subtended by the horizontally placed LEDs was 14.5 degrees. Presentation of the visual signal was unilateral, left or right, whenever heading was out of tolerance by more than one degree. The LED display functioned as a command indicator. In order to terminate it and thus null an error signal, the pilot was required to make a heading correction toward the side on which the visual stimulation was presented. In other words, a horizontal strobe of LEDs on the right side indicated that the pilot should turn the control wheel downward to the right. When the pilot was again within one degree of the desired heading, the visual stimulation was terminated. The indicated heading correction was compatible with the required response.

Procedure

Training

All instruments were explained to the pilots by the experimenter before they got into the cockpit of the simulator. Before flying, the pilots listened to standardized instructions, again explaining all of the instruments. In addition, the instructions outlined the basic maneuvers that would be required of them (climb, straight and level, descent, and right and left turns).

In the flying portion of the training session, the pilots were given a 35 minute flying lesson by the experimenter. During this time, pilots became proficient to criterion on the maneuvers they would be required to perform during the actual experiment. Training was also given on the use of rudders as any power change required a change in rudder pressure in order to maintain heading. The pilots performed two 500 fpm (2.54 m/sec) climbs, one disregarding heading and the other while trying to maintain a heading of 270 degrees. The pilots also performed two straight and level flight segments while attempting to maintain heading and two descents, one disregarding heading and one with heading. Subjects performed their maneuvers between 500 feet and 2500 feet of altitude and were required to maintain an airspeed of 80 mph (35.75 m/sec). The training criterion consisted of performing the maneuvers without deviating more than 15 degrees from the desired heading. All pilots were required to meet the training criterion using the appropriate heading display for their experimental group: Normal (using the compass dial without LED display), or Peripheral (using only peripheral lights to indicate heading).

Experiment

A ten minute experimental session occurred immediately after the pilots' training session. Each pilot was required to perform one climb, one straight and level flight segment, and one descent. Pilots used either the magnetic heading indicator or the peripheral lights (whichever they had received their training with) to remain on heading.

Dependent variables were recorded continuously using a strip chart recorder. These included deviation in degrees from desired compass heading, average errors in vertical velocity, frequency of heading corrections, and frequency of heading overcorrections. Corrections were defined as movements of the simulator from an incorrect heading toward the desired heading of 270 degrees. Overcorrections were defined as swings in heading from one side of the desired heading of 270 degrees to the other side of the desired heading (i.e., 273° to 268°).

Results

Each pilot was given ten minutes to complete the climb, straight and level flight, and the descent. The middle 150 seconds from each of the three flight segments was taken, making a total of 450 seconds being used for test data. Discrete measures of the pilots' ability to maintain a constant vertical velocity and compass heading were taken at five second intervals. Absolute deviations from these standards were then averaged for each pilot. In addition, the total number of heading corrections and overcorrections were tabulated and averaged for each pilot.

A 2 (sex) x 2 (heading indicator) ANOVA was used for each of the four analyses: deviation from desired heading, deviation from desired vertical velocity, frequency of heading corrections, and frequency of heading overcorrections.

In this experiment, variations in sex and the type of heading indicator yielded no significant decrement in control of vertical velocity, deviations in heading, or in average heading corrections per minute. The means of the three non-significant measures are in tables 2, 3, and 4 respectively as listed above. Table 5 presents the means of heading overcorrections per minute.

Table 2
Average Deviations from Desired Vertical Velocity (fpm)

Sex	Indicator	
	Magnetic	Peripheral
Male	118.8	134.3
Female	126.2	123.7

Table 3
Average Deviation from Desired Heading (degrees)

Sex	Indicator	
	Magnetic	Peripheral
Male	6.99	6.70
Female	7.13	7.59

Table 4
Mean Number of Heading Corrections per Minute

Sex	Indicator	
	Magnetic	Peripheral
Male	4.02	4.77
Female	4.54	5.07

Table 5
Mean Number of Heading Overcorrections per Minute

Sex	Indicator	
	Magnetic	Peripheral
Male	2.14	3.09
Female	2.64	3.60

A 2 (sex) x 2 (heading indicator) ANOVA on heading overcorrections yielded a significant main effect for the type of heading indicator ($F=11.97$, $df=1/36$, $p<.005$) and a main effect nearing significance for sex ($F=3.35$, $df=1/36$, $p<.10$). Comparison of the means showed that subjects using the

peripheral heading indicator made significantly more overcorrections than did subjects using the magnetic heading indicator. Also, women made more overcorrections than did men (approached significance).

Discussion

These results support the findings of previous studies. This experiment again determined that significantly more overcorrections were made using a peripheral display, regardless of the sex of the subject. Also as found in previous studies, flight performance as measured by vertical velocity, heading corrections, and deviations in heading did not vary significantly with type of heading indicator.

From the results, it is apparent that peripheral displays are at least as effective as normal displays, and possibly even more effective at indicating errors. This is indicated by the increased number of heading overcorrections when using the peripheral displays. The results of this experiment indicate that peripheral displays are a feasible alternative to foveally placed instruments.

This study determined that the effectiveness of peripheral displays was not impaired by the sex of the operator; women could use the peripheral display as well as men. An important point that should be noted is that the women subjects used in this study had essentially the same motivation to fly as the men. This should also be the case in pilot training or in an actual flight situation; however, it may not be true of the general population. Also, even with the horizontal orientation of the peripheral display, the results of the present study concur with the results of the previous studies which used vertically oriented displays. This would seem to indicate that a horizontal strobe is neither more nor less compatible with required responses than a vertical strobe. Additionally, the results of this experiment indicate that the effectiveness of peripheral displays are essentially the same for novice pilots and for pilots who have soloed an airplane.

Although peripheral lights proved to be no more effective than normal displays in most instances, the significant number of overcorrections made using peripheral lights indicates that the pilots were more responsive to the error signal which in turn indicates increased attention by the pilot. Therefore, peripheral displays could prove to be effective in cueing a pilot for out of tolerance conditions that might be dangerous.

The five simulator experiments and the one tracking task completed to date in this on-going series of experiments have shown a number of significant facts. First, peripheral displays are at least as effective as normal instrument displays. Secondly, the effectiveness of peripheral displays does not degrade under stress. Third, previous studies have shown that peripheral displays using a steady light are more effective than peripheral displays with strobe lights. Fourth, previous studies have also shown that peripheral displays used redundantly with normal displays enhance pilot performance and, as stated previously, could prove to be effective in warning pilots of dangerous conditions. Fifth, women and men perform equally well using either peripheral or normal displays. Finally, experiments to date indicate that the experience level of the pilots does not affect the effectiveness of the peripheral display.

One study that needs to be conducted in this series is to investigate the effectiveness of training pilots for five hours using either a peripheral display or a normal foveal instrument display. So far, all experiments in this series have only used a maximum of one hour of training. It is possible that a longer period of instrument training would separate the performance levels on the two types of displays more definitively.

References

Bermudez, J. M., Harris, D. A., and Schwank, J. C. H., Peripheral Vision and Tracking Performance under Stress. Proceedings of the 23d Annual Meeting of the Human Factors Society, 1979, 402-406.

Bermudez, J. M., Schwank, J. C. H., Longridge, T. M., Smith B. A., and McCloy, T. M., Effect of Peripherally Presented Visual Signals on Pilot Performance during Flight Simulation. Wright-Patterson Air Force Base, Ohio: Aerospace Medical Research Laboratory Technical Report 78-120, June 1979.

Godfrey, M. F., Smith, B. A., and Schwank, J. C. H, Visual Processing: The effects of peripheral heading cues in flight. Proceedings, Psychology in the Department of Defense, 1978, 47-49.

Nataupsky, M., Schwank, J. C. H., Griggs, E. G., McKay K. E., Jr. and Schmidt, S. D. Effect of Pretraining Criterion on Flight Simulator and Secondary Cognitive Task Performance. Proceedings of the 23d Annual Meeting of the Human Factors Society, 1979, 305-309.

Schwank, J. C. H., Bermudez, J. M., Smith, B. A., and Harris, D. A. Pilot performance during flight simulation with peripherally presented visual signals. Proceedings of the 22d Annual Meeting of the Human Factors Society, 1978, 222-226.

Effects of Presentation- and Test-Trial Training on Motor Acquisition and Retention

Joseph D. Hagman
Army Research Institute
Alexandria, VA 22333

Abstract

Examined the relative acquisition and retention effects of three motor task training methods. Three independent groups of governmental employees (N = 15 per group) were trained to perform a simple positioning task using methods which emphasized either alternation of presentation and test trials, repetition of presentation trials, or repetition of test trials during acquisition. Absolute(unsigned) error revealed that acquisition and short-term retention were best after presentation-trial repetition and presentation- and test-trial alternation. Long-term retention, however, was best following repeated test-trial training. It was concluded that testing effectively improves motor retention and that this improvement can be attained by changing the thrust of training from presentation to testing without increasing expenditures in training time, money and personnel.

A long-term goal of the Army is the development of effective training methods for Army-related skills. Of particular interest is the answer to the question of which training methods promote the best skill acquisition and retention. Much of the theoretical information relating to this question originates from the results of basic research experiments. In these experiments, training typically involves the execution of both presentation (p) and test (t) trials. During p-trials, subjects study information to be learned; during t-trials, they attempt to recall it from memory. The number and sequential arrangement of p- and t-trials used during training has depended on the particular training method adopted.

The question of which training method promotes the best acquisition and retention is difficult to answer because relevant theories make conflicting predictions regarding the relative effects of p- and t-trials on learning. Traditional learning theorists view p-trials as having an effect similar to reinforcement (e.g., Adams & Dijkstra, 1966) and predict that training methods which repeat p-trials should be more effective than those which repeat t-trials. Repetition of t-trials reduces the number of reinforcement opportunities, and thus, should retard both acquisition and retention. Contemporary cognitive learning theorists view information processing activities such as retrieval and internal generation of to-be-learned items as important aspects of acquisition and retention (e.g., Bjork, 1975; Dosher & Russo, 1976). Because t-trials provide an opportunity to perform these activities, training methods which repeat t-trials should be more effective than those which repeat p-trials.

The present experiment tested the validity of these conflicting theoretical predictions by comparing the results of three training

methods which differentially emphasized p- and t-trials during the acquisition of a linear positioning motor skill. The general approach was to allow either repetition or alternation of p- and t-trials prior to a given t-trial during training and to compare the relative effects of this manipulation on motor skill acquisition and retention.

METHOD

Subjects

Forty-five governmental employees (27 men and 18 women) volunteered to participate in the experiment.

Apparatus

Movements were performed in a left-to-right fashion on a linear positioning apparatus which consisted of two metal sliding elements attached to two 35 inch (88.9 cm) metal rods. These rods were attached to a metal frame which rested on a standard table top. Subjects used one slide to make both p- and t-trial movements while the experimenter used the other slide to stop movement during p-trials. There was also a chin rest to control head movements and body position; earphones through which subjects heard tape-recorded procedural instructions; and a blindfold to prevent subjects from using visual cues during movement execution.

Design and Procedure

The experiment contained an acquisition and retention segment as shown in Figure 1. Acquisition consisted of 18 training trials divided into three cycles of six trials each. Each cycle contained p- and t-trials. P-trials were experimenter-defined in that subjects contacted a mechanical stop after moving an experimenter-determined distance of 250 mm. T-trials were subject-defined in that subjects were not constrained by the stop in their attempts to reproduce (recall) the criterion distance. The sequence of p- and t-trials within cycles differed for each of three training groups. For Group STANDARD, a cycle consisted of three p- and three t-trials administered in alternation. For Group PRESENTATION, the first five trials in each cycle were p-trials and the sixth was a t-trial. For Group TEST, the first trial was a p-trial and the next five were t-trials. Thus, training was such that a t-trial occurred every sixth trial for all three groups. The retention segment of the experiment consisted of a single t-trial performed at both 3 minutes and 24 hours after trial 18 of acquisition. Each of the total of 20 movements began from a different starting position to prevent subjects from using stopping location as an aid to learning movement distance.

RESULTS AND DISCUSSION

Algebraic (signed) and absolute (unsigned) error scores were recorded for each t-trial performed during the acquisition and the retention segments of the experiment. Scores for each segment were analyzed separately. Only absolute error scores revealed significant effects of interest, and thus, only they are reported.

Acquisition

All three training methods enhanced acquisition performance. Final acquisition performance, however, was best when training involved either p-trial repetition or p- and t-trial alternation rather than t-trial repetition.

Mean absolute error scores for all acquisition t-trials are shown on

the left in Figure 2. Initial statistical analyses were restricted to the scores on those t-trials which coincided temporally for all three training groups, i.e., trials 6, 12 and 18. These scores were analyzed using a Groups (STANDARD, PRESENTATION, TEST) by Trials (6, 12, 18) mixed factorial analysis of variance (ANOVA). The Groups factor was between-Ss and the Trials factor was within-Ss. Significant ($p < .05$) main effects of Trials, F (2,84) = 4.08, and Groups, F (2,42) = 4.48 were found. Individual comparisons revealed that the Trials effect was caused primarily by a decrease in error between trials 6 and 12 and that the decrease between trials 12 and 18 was nonsignificant. The Groups effect showed that the groups which experienced either p-trial repetition (PRESENTATION) or p- and t-trial alternation (STANDARD) during training performed better at the end of each training cycle and at the end of training than the group which experienced t-trial repetition (TEST). Visual inspection of Figure 2 revealed the highly variable nature of the TEST group's acquisition performance. This variability took the form of increased within-cycle error followed by decreased between-cycle error and is consistent with that reported by researchers using verbal tasks (e.g., Izawa, 1970). Within-cycle error increases resulted from interference generated by repeated t-trial movements. Although t-trial movements were not intended to produce interference, they usually were different from the criterion movement, and thus, possessed this capability (Hagman, 1978). The between-cycle error decreases appear to be a function of t-trial repetition potentiating the effect of p-trials. This potentiation effect could have been caused by an increased ability on the part of the subjects to recognize their own recall performance as a function of t-trial repetition. As a result, they were better able to discriminate their movement from the criterion movement and make the appropriate adjustments needed for more accurate recall. Although speculative, the notion that increased discrimination ability results from t-trial repetition is consistent with earlier motor skill research (Newell, 1974).

Retention

The primary benefit of t-trial repetition during training was enhanced long-term retention. It was found that relative to the other two training methods, long-term retention was improved substantially by the training method which allowed opportunities for repeated testing during acquisition.

Mean absolute error scores for retention are shown on the right in Figure 2. Retention was analyzed using a Groups (STANDARD, PRESENTATION, TEST) by Retention Interval (Immediate, 3 minutes, 24 hours) ANOVA. Trial 18 performance was used to indicate immediate recall accuracy. A significant Groups by Retention Interval interaction was found F (4,84) = 4.00. This resulted from an increase in error over time for the STANDARD and PRESENTATION groups an a decrease in error for the TEST group. Individual comparisons of simple effects revealed that at the end of training (trial 18) the TEST group displayed greater error than either the STANDARD or the PRESENTATION group, and that no difference existed between the error of these latter two groups. Thus, group order based upon error magnitude was TEST > PRESENTATION = STANDARD. Three minutes after acquisition training the only change in group performance was an increase in error for the PRESENTATION group. The TEST and PRESENTATION

group scores did not differ 3 minutes after acquisition and their average error score was greater than that of the STANDARD group, i.e., TEST = PRESENTATION > STANDARD. Both the PRESENTATION and the STANDARD group displayed significant error increases between 3 minutes and 24 hours after training while TEST group error decreased. As a result, 24 hours after training TEST group performance was superior to that of the STANDARD group which, in turn, was better than that of the PRESENTATION group. Group order in terms of error magnitude was TEST < STANDARD < PRESENTATION. Thus, emphasis on p-trials during training resulted in rapid and extensive retention losses whereas p- and t-trial alternation enhanced short-term retention and t-trial repetition enhanced long-term retention of motor skill.

The most plausible explanation for the superior long-term retention shown by the TEST group relies on the recent finding that, in general, persons can remember subject-defined movements better than experimenter-defined movements (e.g., Kelso, 1977). If this is true, subjects in the TEST group may have relied on their retention of performance at subject-defined t-trials rather than experimenter-defined p-trials to support long-term retention. Assuming that reliance on a particular movement type is a function of its memorial strength and that strength increases with repetition during training, TEST group subjects should have relied on their retention of t-trial performance and PRESENTATION group subjects should have relied on their retention of p-trial performance. This dependence on different memorial cues produced the different retention performance of the groups. The inability of TEST group subjects to show superior short-term retention was due to the persistence of retroactive interference developed during trials 15-18 of acquisition. Apparently, this interference dissipated over the 24 hour retention interval.

CONCLUSIONS

The results of the present experiment reveal that testing should be viewed as an effective way to enhance long-term retention of motor skills. This enhanced retention can be achieved by changing the emphasis of training from presentation to testing and would occur without the negative aspects of added expenditures in training time, money and personnel. Future research should be directed toward answering the question of whether or not these benefits will generalize to other types of motor skills. Of particular interest should be the examination of procedural motor skills which require the execution of successive movements in the correct serial fashion and are characteristic of many real-life Army tasks.

REFERENCES

Adams, J. A., & Dijkstra, S. Short-term memory for motor responses. Journal of Experimental Psychology, 1966, 71, 314-318.

Bjork, R. A. Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. L. Solso (Ed.), Information processing and cognition the Loyola symposium, Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers, 1975.

Dosher, B. A., & Russo, J. E. Memory for internally generated stimuli. Journal of Experimental Psychology: Human Learning and Memory, 1976, 2, 633-640.

Hagman, J. D. Specific-cue effects of interpolated movements on distance and location retention in short-term motor memory. Memory and Cognition, 1978, 6, 432-437.

Izawa, C. Optimal potentiating effects and forgetting-prevention effects of tests in paired-associate learning. Journal of Experimental Psychology, 1970, 83, 340-344.

Kelso, J. A. S. Motor control mechanisms underlying human movement reproduction. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 529-543.

Newell, K. M. Knowledge of results and motor learning. Journal of Motor Behavior, 1974, 6, 235-244.

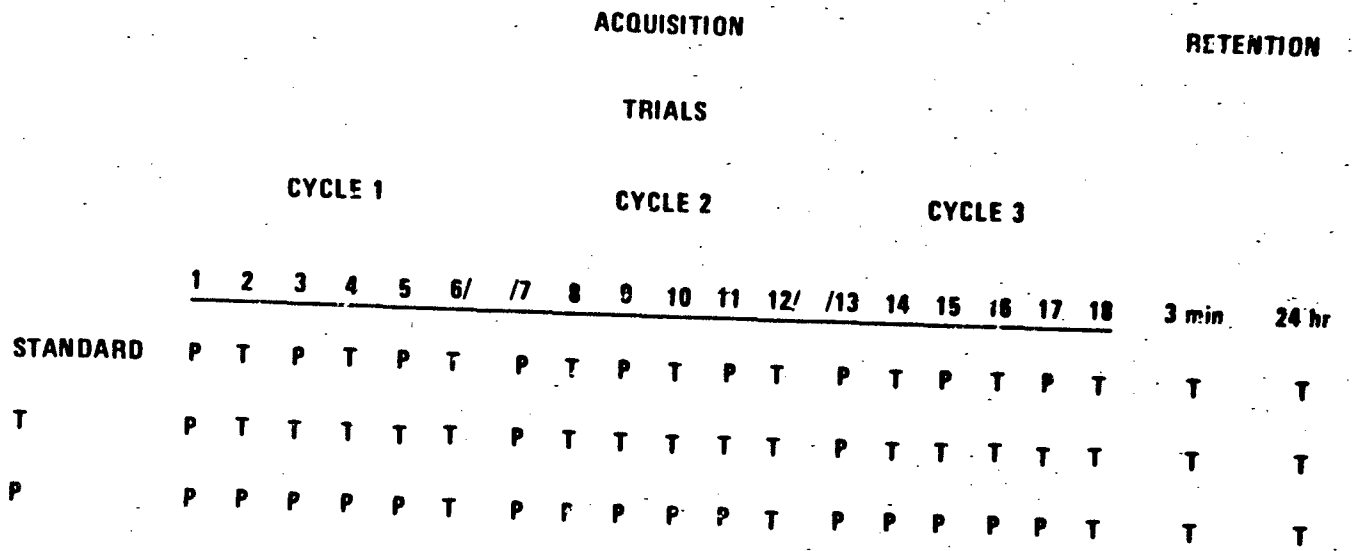


Figure 1. Trial sequence of each training group at acquisition and retention (P=Presentation; T=Test).

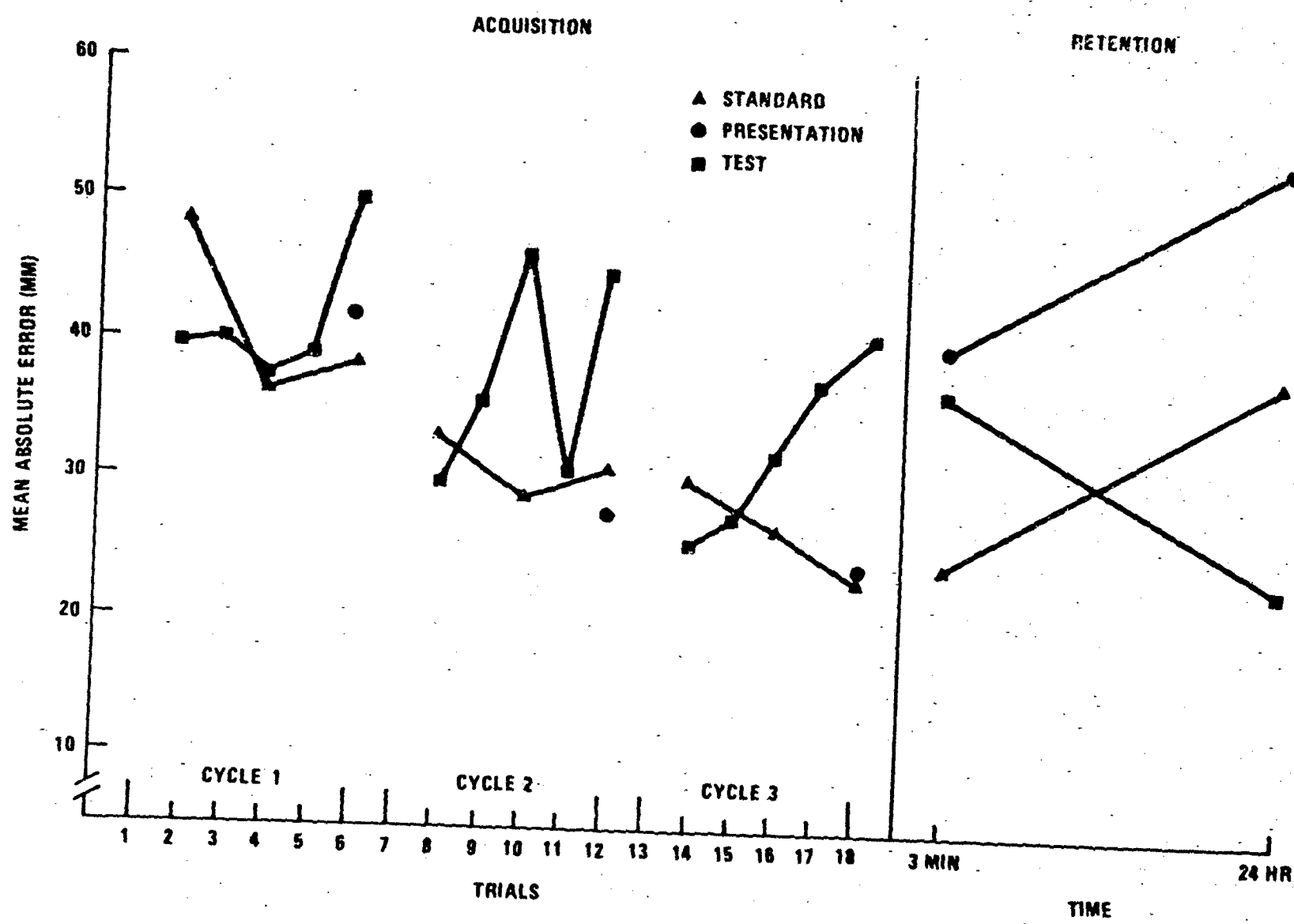


Figure 2. Mean absolute error resulting from t-trial execution by the STANDARD, PRESENTATION and TEST training groups at acquisition and retention.

Comparison of Males and Females in the Prediction of Basic Flight Performance

Jefferson M. Koonce and Gene A. Berry
United States Air Force Academy, Colorado

Abstract

Fifty-one male and fifty-two female subjects were given a battery of cognitive tests and a battery of psychomotor test in an attempt to determine wnether there are sex differences in the prediction of basic flight performance on a desk-top flight trainer. The resultant multiple regression equations for males had different significant predictors than those for females. Possible implications for pilot training candidate selection programs are discussed.

For many years paper and pencil tests have been used by the military services in the screening of candidates for most officer training and flight training programs. Gould (1978) reviews the development of the Air Force Officers Qualifying Test (AFOQT) which has been used by the U. S. Air Force since 1951 and has undergone fourteen revisions in the past twenty-seven years. The test's results are organized into five composite scores (officer quality, verbal, quantitative, pilot, and navigator-technical) based upon various combinations of the eighteen subtests which include electrical maze, rotated blocks, and aerial landmarks.

Psychomotor ability has been used as a predictor of pilot training success off and on since 1919 (Dockeray, 1921). The use of the psychomotor tests was discontinued in the U. S. Air Force in the early 1950s because of difficulties in calibrating and maintaining the machines under decentralized testing conditions. Passey and McClaurin (1966) provide a comprehensive review article on psychomotor testing over the years and suggest certain refinements for the updating of psychomotor test equipment. McGrevy and Valentine (1974) report on the development and validation of two psychomotor tasks presented by a solid state psychomotor device (PM) which is reliable, easily maintained, and provides accurate measures of subject performance. The two tasks are presented and scored automatically with a cassette tape giving the subject the instructions on how to perform the tasks and providing the cures for the presentation and scoring circuits. The first task is a two-hand coordination pursuit task, and the second task is a complex coordination test requiring x-y compensatory tracking with a joystick and a single z-axis tracking using footpedals. Both tasks are fixed in difficulty and the device records x and y deviation on the first task and x, y, and z deviation on the second task. McGrevy and Valentine state (pg. 9) that "the addition of the psychomotor tests to the AFOQT can enhance the prediction of pilot training success."

Recently, Savage, Williges, and Williges (1978a, 1978b) have noted that in predicting performance on a psychomotor task a larger proportion of criterion variance can generally be accounted for when the regression equations are individually tailored by sex as opposed to a single overall regression equation. The predictor variables that they used were a pursuit rotor task and five paper and pencil tests as follows: Identical Pictures Test (perceptual speed), Map Memory Test (visual memory), Cube Comparison

Test (spatial orientation), Maze Tracing Test (spatial scanning), and the Embedded Figures Test (field independence). The criterion was time-to-exit on a fixed difficulty or adaptive two-axis pursuit tracking task. For the fixed difficulty time-to-exit, the identical pictures test and the embedded figures test were the significant predictors for the males, and the embedded figures test and the map memory test were the significant predictors for the females. Because a combination of the AFOQT performance and the PM device performance has been proposed to be used in the screening of pilot training candidates (McGrevy and Valentine, 1978), this study will attempt to determine whether there is a sex difference in predicting flight training performance of simple maneuvers in a simulator using as predictors the battery of Savage, Williges, and Williges which has some subtests which are similar to those in the AFOQT and the same PM device proposed by McGrevy and Valentine.

Method

Subjects. Fifty-one male and fifty-two female volunteer fourth class cadets at the U. S. Air Force Academy participated in three 50-minute sessions over a two month period.

Procedure

Session 1. The first session consisted of small group administration of four paper and pencil tests. The tests are those used by Savage, Williges, and Williges (1978) and were selected because of their similarity to subtests forming the pilot scale of the AFOQT. The tests and cognitive-perceptual areas tapped are as follows:

IP - Identical Pictures Test (perceptual speed)
MM - Map Memory Test (visual memory)
CC - Cube Comparison Test (spatial orientation)
MT - Maze Tracing Test (spatial scanning)

Session 2. Each subject was given the individual version of the Embedded Figures Test (EFT) (Witkin, Oltman, Raskin and Karp, 1971) to assess field independence. The score was the total seconds to locate the twelve figures. Then they were given a conventional pursuit rotor (PR) Test with their score being the average time on target over six thirty-second trials at 60 rpm with ten seconds between trials. Finally, each subject performed two psychomotor tasks on the automated device (PTD) developed by McGrevy and Valentine (1978): a two-hand coordination pursuit task (PTD1) ans a stick and rudder complex coordination task (PTD2). These tasks (PTD 1&2) were performed on a device previously recommended as a pilot aptitude selection device. The errors on the psychomotor device PTD1 were summed x and y deviation, and they were combined to form a resultant vector error score. The PTD2 test resulted in x, y, and z error scores which were combined as rudder pedal error (z) plus the x-y vector error from the stick and expressed as PTD2E.

Session 3. During the final session each subject was instructed in how to perform basic flight maneuvers using a ATC desk-top trainer. The subject was then given practice on four maneuvers: climb, cruise, descent, and standard rate turn. Finally, each subject was tested on his or her performance while flying these four maneuvers in simulated smooth and turbulent air conditions. The subjects' errors were recorded by the experimenters in booklets designed for those tasks, similar to the Pilot Performance Record (Koonce, 1979). The errors were standardized across all

subjects and averaged per maneuver and summed for smooth air maneuvers, rough air maneuvers, and all maneuvers combined.

Results and Discussion

The performance of the males and females on the predictor tests were in the directions anticipated in that the females were somewhat faster on the IP test, while the males tended to be better on the MM, CC, and MT tests. Only on the CC test was the difference statistically significant ($p < 0.001$) Also, the males were faster on the EFT, but not significantly so. On all of the psychomotor tests and the simulator tests in smooth air, rough air, and overall, the males were significantly better than the females ($p<0.001$)

The lack of significant differences between the sexes on the cognitive tests might be due in part to the selection process for cadets at the Academy which tends to result in a cognitively more homogeneous group than would be expected in the general population.

The scores for the cognitive and psychomotor variables were used to generate multiple regression (stepwise) equations for each criterion variable: performance in smooth air, performance in rough air, and overall flight performance. All of the correlations are significant ($p < 0.05$).

Multiple Correlations and Significant Predictors of Basic Flight Maneuvers

		R
Smooth Air		
Males	TZS = 2.19 + 4.68 PTD2E	.485
Females	= 4.95 - 0.142 PR - 1.55 CC	.374
Both	= 3.44 + 253.1 PTD2E - 1.008 PR	.460
Rough Air		
Males	TZR = 3.98 - 0.280 IP + 203.2 PTD2E	.428
Females	= 4.76 - 0.115 PR - 2.427 CC	.388
Both	= 3.17 + 0.614 SX - 1.528 PR - 3.15 CC	.520
All Maneuvers		
Males	TZ = 6.66 + 1.304 PTD2E - 2.049 IP	.497
Females	= 9.71 - 0.257 PR - 0.106 CC	.409
Both	= 5.44 - 1.109 SX - 0.152 PR + 275PTD2E	.526

Over the entire sample without using sex (SX) as a predictor variable, the prediction of simulator performance was not significantly different from those correlations for either sex seperately.

For the males, the PTD2E and the IP tests were the significant predictors; but for the females the PR and CC were the significant predictors. The great dissimilarities in the equations for males and females tends to support the notion that to select those females with the greatest probability of success in pilot training one should develop the weights for the predictor tests from females in the population as opposed to those weights derived from the males in the population.

References

Dockeray,F.C. & Isaacs,S. Psychological research in aviation in Italy, France, Enagland, and the A.E.F. Journal of Comparative Psychology, 1921, 1, 115-148.

Ekstrom,R.B., French,J.W., Harman,H.H., & Dermen,D. Manual for kit of factor-referenced cognitive tests. Princeton, N.J.: Educational Testing Service, 1976.

Gould,R.B. Air Force officer qualifying test form N: Development and standardization. Brooks Air Force Base, Texas: Air Force Human Resources Laboratory, Technical Report AFHRL-TR-78-043, August 1978.

Koonce,J.M. Predictive validity of flight simulators as a function of simulator motion. Human Factors, 1979, 21, 215-223.

McGrevy,D.F. & Valentine,L.D. Validation of two aircrew psychomotor tests. Lackland Air Force Base, Texas: Air Force Human Resources Laboratory, Technical Report AFHRL-TR-74-4, January 1974.

Passey,G.E. & McLaurin,W.A. Perceptual-psychomotor tests in aircrew selection: Historical review and advanced concepts. Lackland Air Force Base, Texas: Personnel Research Laboratory, Technical Report PRL-TR-66-4, June 1966. (AD 636 606)

Savage,R.E., Williges,R.C., & Williges,B.H. Individual differences in motor skill training. Proceedings of the Sixth Psychology in the DoD Symposium, 1978a.

Savage,R.E., Williges,R.C., & Williges,B.H. Cross-validation of regression equations to predict performance in a pursuit tracking task. Proceeding of the Human Factors Society - 22nd Annual Meeting, 1978b.

Witkin,H.A., Oltman,P.N., Raskin,E., & Karp,S.A. A manual for the embedded figures test. Palo Alto, California: Consulting Psychologists Press, 1971.

Acknowledgments

Contractual support for this study was provided by the Life Sciences Program, Air Force Office of Scientific Research, Project Order AFOSR-PO-79-0009. Major [illegible] A. Thorpe was the scientific monitor of the project order. Also, we wish to acknowledge the assistance of Richard J. Becker of Virginia Polytechnic Institute and State University and Charles A. Beaver of the U. S. Air Force Academy in the collection of data.

GENDER DIFFERENCES IN AIRCRAFT SIMULATOR PERFORMANCE

Thomas M. McCloy
Department of Behavioral Sciences and Leadership
United States Air Force Academy

Abstract

In this study, equal numbers of males (28) and females (28) flew three basic instrument maneuvers in a GAT-1 aircraft simulator. Males performed better than females on all performance variables associated with the three maneuvers. The results suggest the need for further research to determine the source of the differential performance.

The need for more information regarding male and female differences in psychomotor performance was emphasized in a recent review by Hudgens and Billingsley (1978). Although data exist investigating simple motor skill performance differences (Singer, 1975), relatively little is known involving complex task performance. This is true despite the fact that women are becoming increasingly more involved in occupations heretofore reserved for men.

The present study was designed to investigate male and female performance on a complex psychomotor task -- flying an aircraft simulator. This task is of particular importance because women have recently been included in military flight training programs. Should sex differences be found in simulator performance, it may have implications for selection, training, and evaluation procedures for potential male and female pilots.

Method

Subjects

Fifty-six Air Force Academy cadets, 28 male and 28 female, participated in the experiment. They were volunteers from the freshman, sophomore and junior classes. None of the participants had previous private pilot or aircraft simulator performance. All subjects had at least 20-20 correctable vision.

Apparatus.

Subjects were trained and tested in a Singer-Link GAT-1 Simulator. The performance characteristics of the GAT-1 are similar to those of a single-engine, propeller driven aircraft (e.g. a Cessna 172). The instruments and controls correspond to those typically installed in that type of aircraft.

Procedure

Training, practice, and testing for all subjects lasted approximately 50 minutes. For the first 35 minutes, subjects received instructions explaining the instruments and controls utilized in flying the simulator, and practiced basic instrument maneuvers. At the end of the instruction and practice period, the subjects were tested on three maneuvers: 1) a constant heading, constant rate climb of 2,000 feet; 2) straight and level flight for 2 minutes; and 3) a constant heading, constant rate descent of 2,000 feet. Analog signals for heading and vertical velocity (rate) were recorded for performance evaluation.

Results and Discussion

Subjects' heading and vertical velocity (VVI) scores for the three different maneuvers provided the basis for simulator performance analyses. Both heading and vertical velocity represented deviations from desired values. Consequently, lower scores were indicative of better performance. Table 1 depicts the 2 by 3 matrix that represents each subject's scores.

Table 1 Matrix representing subject's simulator scores.

	Climb	Straight and Level	Descent
Heading	S-CLIMBH	S-SANDLH	S-DESH
VVI	S-CLIMBV	S-SANDLV	S-DESV

Note. VVI - vertical velocity

Prior to statistical analyses, cell scores were standardized to a mean = 0 and a standard deviation = 1 (the variables representing the standardized cell values are depicted in Table 1). Standardized heading and VVI scores were then added to provide overall scores for each of the three maneuvers. Finally, an overall total performance score was calculated by adding scores from each of the six cells. The overall total score, the three different maneuver scores, and the scores representing general heading and vertical velocity control were evaluated as indicators of simulator performance (see Table 2).

Table 2 t-tests for gender differences on the six simulator performance scores.

VARIABLE	GENDER	SCORE	t-TESTS
STANTOT	Female Male	1.443 -1.443	$t = (54) = 3.09$, $p < .003$
STANHEAD	Female Male	.692 -.692	$t = (54) = 2.08$, $p < .04$
STANVVI	Female Male	.750 -.750	$t = (54) = 3.53$, $p < .001$
STANCLIM	Female Male	.529 -.529	$t = (54) = 2.33$, $p < .02$
STANSL	Female Male	.436 -.436	$t = (54) = 2.264$, $p < .02$
STANDES	Female MALE	.478 -.478	$t = (54) = 2.196$, $p < .03$

Note. STANTOT = S-CLIMBH+S-CLIMBV+S-SANDLH+S-SANDLV+S-DESH+S-DESV

STANHEAD = S-CLIMBH+S-SANDLH+S-DESH

STANVVI = S-CLIMBV+S-SANDLV+S-DESV

STANCLIM = S-CLIMBH+S-CLIMB-V

STANSL = S-SANDLH+S-SANDLV

STANDES = S-DESH+S-DESV

Although males performed significantly better than females on all six simulator scores (as indicated in Table 2), an examination, by gender, of the intercorrelations of the six simulator variables (Tables 3 & 4) suggest the pattern of responding was very similar for both males and females. The first relationship of interest is that between heading and vertical velocity. Since these were the two performance variables it is conceivable that different strategies may have been employed, perhaps maximizing one score at the expense of the other. In general, this does not appear to be the case, as both males and females exhibit significant and similar correlations between heading and vertical velocity performance. This suggests there was more of a tendency for subjects to either do well on both, or poorly on both, as opposed to emphasizing one over the other.

The relationship of the three separate maneuvers to each other is also of interest. Although the performance measures for all three maneuvers involved deviations from desired headings and vertical velocities, there were several basic differences between the maneuvers that might help explain performance differences. The GAT is designed to closely simulate light aircraft handling characteristics. Consequently, it generally requires some right rudder in a climb to offset the effects of rigging and asymmetrical loading of the propeller, and left rudder in a descent to compensate for rigging. Straight and level flight, on the other hand, does not normally require the use of rudder. Since the use of the rudder necessitates the checking of another instrument (the turn and slip indicator) as well as the operation of another control, it may be that climbs and descents are more related and perhaps more difficult than straight and level. The pattern of correlations for both males and females suggest the likelihood of this interpretation. Both males and females exhibit significant correlations between climb and descent performance, and nonsignificant correlations between straight and level and either climbs or descents. Additionally, the relationship between the measure of overall performance, STANTOT, and climb and descent performances tended to be stronger than that indicated between STANTOT and straight and level. This finding was the same for both males and females.

Table 3 Intercorrelations of the six simulator performance variables for females. Significance levels are included in parentheses.

	STANTOT	STANHEAD	STANVVI	STANCLIM	STANSL	STANDES
STANTOT	-					
STANHEAD	.895 (.0001)	-				
STANVVI	.810 (.0001)	.464 (.0130)	-			
STANCLIM	.690 (.0001)	.562 (.0019)	.631 (.0003)	-		
STANSL	.584 (.0011)	.508 (.0058)	.492 (.0078)	-.032 (.8702)	-	
STANDES	.846 (.0001)	.827 (.0001)	.594 (.0009)	.438 (.0196)	.344 (.0734)	-

Table 4 Intercorrelations of the six simulator performance variables for males. Significance levels are included in parentheses.

	STANTOT	STANHEAD	STANVVI	STANCLIM	STANSL	STANDES
STANTOT	-					
STANHEAD	.936 (.0001)	-				
STANVVI	.716 (.0001)	.424 (.025)	-			
STANCLIM	.790 (.0001)	.727 (.0001)	.591 (.0009)	-		
STANSL	.657 (.0001)	.724 (.0001)	.254 (.193)	.236 (.2272)	-	
STANDES	.819 (.0001)	.687 (.0001)	.745 (.0001)	.487 (.0087)	.356 (.0633)	

Conclusions

The finding of significant sex differences indicates a need for further investigation to determine the source of the differential simulator performance. Additionally, current pilot selection, training and evaluation instruments that are based on male performance standards should be re-evaluated to determine their appropriateness when applied to females.

References

Hudgens, G.A. and Billingsley, P.A. Sex: The missing variable in human factors research. Human Factors, 1978, 20, 245-249.

Singer, R.N. Motor learning and human performance. New York: Macmillan, 1975.

The Use of Personality Characteristics as Predictors of Psychomotor Performance

G. A. Berry, D. A. Harris and J. M. Koonce

United States Air Force Academy

Abstract

Canonical correlation analyses were performed on 17 cognitive, perceptual, motor, demographic and performance measures for 98 subjects. The addition of personality trait scores significantly improved linear predictor-performance criterion correlation and increased variance accounted for (R^2) by 27%. Results indicate the inclusion of relevant personality data may improve prediction of complex task performance.

Perhaps no area of applied psychology has received more sustained interest than the attempt to reliably predict performance on highly complex tasks from simpler, more attainable measures. This effort has undoubtedly been motivated by pressures of time, money and sometimes human life savings which could result from accurate prediction of training and ultimate performance of complex tasks such as flying an aircraft.

Researchers in the area have successively added different types of predictor variables. From early attempts based on simple motor tasks such as the pursuit rotor, other types of variables have been added. These have included perceptual measures, paper-and-pencil cognitive tests, and more recently complex integrations of these types of measures such as complex perceptual-motor devices such as the Psychomotor Test Device of Systems Research Laboratories (McGrevy & Valentine, 1974) to predict flight training performance.

Another set of measures which might contribute to prediction is comprised of measures of various personality traits. This possibility served as the genesis of this study.

Efforts to predict complex performance from simpler, more attainable measures have typically used multivariate statistical techniques. Some have relied on factor analysis to tap sets of highly interrelated measures, demographic categories, etc. More typically they have attempted to develop multiple regression equations with weights assigned each contributing predictor.

A third multivariate approach, canonical correlation analysis, is useful when seeking the relation between two sets of variables (Tatsuoka, 1971). In performance prediction, these could represent predictor variables and criterion variables. Flight training performance has typically been measured as scores on a number of variables rather than a single overall criterion. This technique seems more appropriate than developing factors through factor analysis comprised of confusing mixes of predictor and criterion variables. Likewise it appears better than either developing multiple regression equations for each performance measure or a single multiple regression equation to predict some questionable overall summation performance measure on a task so complex that there really is none. Application of canonical correlation analysis will allow predictors and criterion, or performance, measures to be

kept separate, while permitting a single overall analysis.

In general, canonical correlation analysis addresses predictor-criterion variable set relations by determining linear combinations of predictor scores that are most highly correlated with linear combinations of criterion measures. To establish that additional predictor variables or sets of variables can add to the prediction of performance (as represented by a set of scores) the predictors can be successively added and the resulting relations compared. In this way the value of utilizing the additional variable or set of variables can be assessed. Thus a decision could then be made with regard to possible gain in prediction versus the cost in terms of time, effort, money, danger, inconvenience, etc. of obtaining scores on the additional predictor variables under consideration.

The purpose of the current study was to assess, in a general manner, whether the addition of personality trait measures could add to the prediction of a set of performance measures. The previously discussed canonical correlation approach was selected to measure the value of using these measures when added to the more established motor, demographic, perceptual and cognitive predictors. This analysis would permit a more general assessment than immediately attempting to develop regression equations to predict specific performances. It seemed particularly appropriate given the historic difficulty in establishing reliable multiple regression prediction equations which can survive validation tests for various tasks, other groups, both sexes, etc. even without the inclusion of personality measures (Becker, Williges, Williges & Koonce, 1979). This study then was designed to provide a general look at the possible value of including a set of personality trait measures in future attempts to predict complex skill performance.

Method

Subjects

Ninety-eight cadet undergraduate volunteers of the U.S. Air Force Academy, 49 of each sex, participated in the study.

Procedure

During 1979, all subjects were measured on a number of tasks, tests, and demographic characteristics. Paper-and-pencil cognitive measures from the battery of Ekstrom, French, Harmon and Derman (1976) and what they purport to measure were as follows: (1) Identical Pictures Test (perceptual speed), (2) Map Memory Test (visual memory), (3) Cube Comparisons Test (spatial orientation), and (4) Maze Tracing Test (spatial scanning). In addition, each subject was administered the perceptual measure, the Embedded Figures Test for field dependence-independence (Witkin, Oltman, Raskin & Karp, 1971). Performance measures were taken on: (1 and 2) Test Device Tests 1 and 2 of the System Research Laboratories Psychomotor Test Device (McGrevy & Valentine, 1974), (3) pursuit rotor and (4) a performance score on the ATC desk top aircraft trainer.

In addition, the following personality measures from the Work and Family Orientation Questionnaire (Helmreich & Spence, 1978) were used: (1) Mastery, (2) Competitiveness, (3) Work Orientation and (4) Personal Unconcern. From the Personal Attributes Questionnaire (Spence, Helmreich & Stapp, 1974) scores were gathered on the traits of (1) Masculinity, (2) Femininity and (3) Masculinity-Femininity. Descriptions of all traits and their measurement devices are presented in the references.

Analysis

Canonical correlation analysis was performed in two stages using the CANCORR subprogram of the Statistical Package for the Social Sciences (Nie, Hull, Jenkins, Steinbrenner & Bent, 1975).

Stage 1. A canonical correlation analysis was performed using sets presented below. Set 1 represents a combination of cognitive and perceptual predictor variables. Set 2 is a set of performance variables.

Set 1	Set 2
IP - Identical Pictures Test	PR - Pursuit Rotor
MM - Map Memory Test	PTD1 } Psychomotor Test
CC - Cube Comparisons Test	PTD2 } Device Tasks 1 & 2
MT - Maze Tracing Test	TNR - A/C Trainer
EFT - Embedded Figures Test	

Stage 2. For the second analysis, the following personality trait measures were added to Set 1:

M - Masculinity
F - Femininity
MF - Masculinity-Femininity
W - Work Orientation
MA - Mastery
C - Competitiveness
PU - Personal Unconcern
SX - Sex

Results

The results of the first canonical correlation analysis are presented in Table I. As can be seen, one canonical variate was significant. In

TABLE I
CANONICAL CORRELATION ANALYSIS BETWEEN COGNITIVE/ PERCEPTUAL VARIABLES AND PERFORMANCE VARIABLES

Number	Eigenvalue	Canonical Correlation	Chi-Square	d.f.	Significance
1	.160	.400	31.586	20	0.048

Coefficients for Perceptual/Cognitive Variables

VARIABLE	COEFFICIENT
IP	-.045
MM	.231
CC	.557
MT	.593
EFT	-.141

Coefficients for Performance Variables

VARIABLE	COEFFICIENT
PR	.177
PTD1	-.210
PTD2	-.555
TNR	-.387

addition the linear combinations were correlated .400. The perceptual/ cognitive variables which contributed the most to the correlation were

the cube comparison (CC) and maze tracing (MT) tests. The performance variables which contributed most to the canonical correlation were the psychomotor test device-test two (PTD2) and the ATC desk top aircraft trainer (TNR) performance.

Table II reflects the results of the second canonical correlation analysis. In this analysis the seven personality trait measures and gender were added to the variables used in the first analysis.

TABLE II
CANONICAL CORRELATION ANALYSIS BETWEEN COGNITIVE/ PERCEPTUAL, PERSONALITY AND PERFORMANCE VARIABLES

Number	Eigenvalue	Canonical Correlation	Chi-Square	d.f.	Significance
1	.430	.656	91.420	52	.001

Coefficients for Perceptual/Cognitive, Personality Variables		Coefficients for Performance Variables	
VARIABLE	COEFFICIENT	VARIABLE	COEFFICIENT
M	.363	PR	.174
F	.341	PTD1	-.292
MF	.201	PTD2	-.572
W	-.148	TNR	-.296
MA	-.056		
C	.026		
PU	-.086		
SX	-.668		
IP	.203		
MM	.082		
CC	.092		
MT	.261		
EFT	.032		

Again, only one canonical variate was significant. However, by adding the personality trait variables the amount of variance accounted for by the canonical correlation increases 27%. Variables which contributed the most to the canonical correlation were: (1) gender (SX), (2) masculinity (M), (3) femininity (F), and (4) the second test on the psychomotor device (PTD2).

Conclusions

The major finding of this investigation was that personality variables substantially increase the correlation between linear combinations of performance variables and cognitive/perceptual, personality variables. Thus, the inclusion of various relevant personality variables may increase the predictability of complex psychomotor task performance.

References

Becker, R. J., Williges, B. H., Williges, R. C., & Koonce, J. M. Prediction of performance in motor skills training. Proceedings of the Human

Factors Society - 23rd Annual Meeting, 1979, 300.

Ekstrom, R. B., French, J. W., Harman, H. H., & Derman, D. Manual for kit of factor-referenced tests. Princeton, N.J.: Educational Testing Service, 1976.

Helmreich, R. L., & Spence, J. T. The Work and Family Orientation Questionnaire: An objective instrument to assess components of achievement motivation and attitudes toward family and career. JSAS Catalog of Selected Documents in Psychology, 1978, 8, 35 (MS #1677).

McGrevy, D. F., & Valentine, L. D. Validation of two aircrew psychomotor tasks (AFHRL Tech. Rep. 74-4). Lackland Air Force Base, TX: Air Force Human Resources Laboratory, January 1974.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. Statistical package for the social sciences (2nd Ed.). New York: McGraw-Hill, 1975.

Spence, J. T., Helmreich, R., & Stapp, J. The Personal Attributes Questionnaire: A measure of sex-role stereotypes and masculinity-femininity. JSAS Catalog of Selected Documents in Psychology, 1974, 4, 43-44, (MS #617).

Tatsuoka, M. M. Multivariate analysis. New York: Wiley, 1971.

Witkin, H. A., Oltman, P. K., Raskin, E., & Karp, S. A. A manual for the Embedded Figures Tests. Palo Alto: Consulting Psychologists Press, 1971.

Training Criteria: The Missing Link in Flight Simulation Research?[1]

Mark Nataupsky, John M. Bermudez, and Valentin W. Tirman

United States Air Force Academy

For the past few years, a research group within the Department of Behavioral Sciences and Leadership has been conducting a series of experiments designed to study characteristics of human information processing in vision and attention during simulated flight. Some of these studies have been aimed at improving pilot performance by means of alternate flight control monitoring devices such as displays placed in the peripheral fields of ambient vision. Other studies have been aimed at investigating the training applications of these alternate forms of displays (e.g., Bermudez, Schwank, Longridge, Smith, and McCloy, 1979; Schwank, Bermudez, Smith, and Harris, 1979).

The present study is related to both of these research aims. In both lines of research, our group has strived to obtain both system and operator performance data under conditions that approximate relevant information processing demands and realistic motor task demands. For example, all our studies have required subjects to perform demanding mental activity during either the difficult vertical S aerial manuever or during an instrument landing approach. We have employed probed memory techniques, modified after the Sternberg technique to impose stress while the pilot flew the simulator. Although these earlier studies have shown memory probes are very demanding, the type of probe used seemed much less difficult than the types of processing demands and probes that typically confront the pilot. Previous memory probes have required little more than recognition memory for locations or for identity of a stimulus or for both (e.g., the location of dials and the identities of readings). In the present experiment, the secondary task included additional processing demands for the pilot to mentally manipulate and to transform data in addition to storing the result for later retrieval. It was assumed that the increased memory loads would interfere with the primary task in a linear fashion.

Additionally, there is evidence, outside of the research literature of flight simulation (e.g., Cofer, 1969; Weitz, 1961) to support the idea that the particular performance criterion level established in training directly affects both the later experimental performance of subjects and the conclusions that the research worker can draw from the performance data. Curiously, we could not locate any studies which used training criteria as an independent variable to assess the strength of such effects. Therefore, we compared the effects of two levels of training criterion against two kinds of visual displays in a flight simulator.

[1]This research was sponsored by Frank J. Seiler Research Laboratory work unit 2303-F1-54 and Aerospace Medical Division Contract AMD/RDO 78-1, Amendment 1.

PRECEDING PAGE BLANK-NOT FI

Method

Design and Subjects

The design was a 2 x 2 x (5) mixed factorial with repeated measures on the last factor. There were two levels of compass heading error tolerance (5 versus 10 degrees) and there were two types of compass heading error indicators (typical dial compass versus light display). There were five levels of memory load (a no-load control, and four loads involving increasing amounts of information to process). The subjects were male, novice Air Force Academy cadet pilots who had completed between nine and 21 hours in a T-41 aircraft. The cadets were randomly assigned, seven to each of the four experimental cells, in order of their appearance for the experiment.

Apparatus. Training and experimental sessions were conducted using a Link-Group General Precisions Systems Incorporated General Aviation Trainer-1 (Gat-1). This flight simulator was modified by mounting a small video screen above the instrument panel to enable presentation of the secondary task memory load. Cockpit design and simulated flight characteristics were highly similar to the T-41 aircraft. Data was recorded using a six channel Brush Stripchart Recorder.

Heading Indicators. Two groups of subjects used the typical T-41 dial compass which was mounted on the instrument panel. The remaining two groups used only a head-mounted system of peripheral vision signals with the dial compass masked from view. The LED signals consisted of five light-emitting diodes (LEDs), green in color, wired in series. Each LED was 2.4 mm in diameter with the LEDs placed six mm apart. The displays provided retinal projections at 55 degrees of arc subtended from the front, center of the person's head. Onset of the light occurred only when the pilot exceeded the desired heading by ± one degree. In order to correct an out-of-tolerance condition, the pilot's correct response was to turn towards the light and null the error.

Procedure

Training

All pilots were trained to criterion on a vertical S aerial manuever for one hour. This maneuver consists of ascending to an altitude of 2000 ft (656.17 m), then alternately, ascending and descending 250 ft (82.02 m) above and below the baseline. This maneuver had to be performed while holding to a heading of 270 degrees, airspeed of 80 mph (35.76 m/sec), and vertical climb and descent at 500 ft per minute (2.54 m/sec). This maneuver is very demanding and requires frequent motor activity and instrument scanning and cross-checking.

For two of the experimental groups, the criterion consisted of performing the vertical S without deviating more than 10 degrees from the desired heading. For the remaining two groups, the error tolerance was five degrees. Thus, there were four experimental groups: 5-DIAL, 10-DIAL, 5-LED, and 10-LED. An instructor pilot monitored performance during training and supplied prompts whenever heading deviations exceeded either the five or 10 degree training criterion.

Experiment

Experimental sessions occurred on the second day following training. Each subject was allowed five minutes of practice (two trials) after which four trials of the vertical S were performed. The subjects all performed a

secondary task involving responding to 60 sets of visual stimuli and memory probes. The duration of each stimulus display was three seconds with a three second interstimulus delay.

Each visual presentation consisted of one of five types of video displays, providing five levels of memory load: (1) zero or no load control, (2) memorizing that a display had blanks, (3) memorizing one letter and the sum of one number, (4) memorizing one letter and the sum of 3 numbers, and (5) memorizing one letter and the sum of 5 numbers. At each level of memory load, the cadets engaged in three cognitive activities: 1) localization, 2) identity, and 3) transformation. The purpose of imposing the memory loads was to approximate levels of cognitive demands analogous to the information processing demands of scanning and cross-checking instrument displays.

Dependent Measures.

(1) Compass Heading errors. These errors were scored as absolute deviations from 270 degrees, sampled every five seconds during each two minute trial.

(2) Compass heading overcorrections. These errors were scored as ratios of the number of times the pilot crossed from one side of the desired heading (270°) to the other side (e.g., 268° to 273°) per trial.

(3) Airspeed errors. These errors were scored as absolute deviations from 80 mph (35.76 m/sec) sampled every five seconds.

(4) Rate of Ascent/descent errors. These errors were scored as absolute deviations from the required 500 ft/min (2.54 m/sec) sampled every five seconds after the attitude was established.

(5) Error Time. This score consisted of the proportion of time the pilot erred by exceeding the five or 10 degree criterion during the experiment.

(6) Memory load scores. These scores consisted of the percent of correct responses to the 60 memory probes in the secondary task.

Results and Discussion

Table 1 presents the means and standard deviations for all measures of performance, at each level of memory load. A 2 (heading error tolerance) x 2 (compass type) x (5) (memory load) mixed analysis of variance on these data yielded significance for the memory load main effect, as follows: (a) compass heading errors, $F=8.25$, $p < .01$; (b) heading overcorrections, $F=15.30$, $p < .01$; (c) airspeed errors, $F=4.92$, $p < .01$ (all df = 4,96); (d) Error Time, $F(1, 15) = 8.26$, $p < .05$; and memory load scores, $F(3,66)= 14.19$, $p < .01$. The ascent/descent rate was not significant. Separate 2-tailed t-tests were calculated between the control condition and each of the four memory load levels, and between the load levels only. These results are also outlined in table 1.

The ANOVA also yielded significance for several important interactions between the type of compass heading indicator and the error tolerance training criterion used in training. These results are presented in table 2.

These results clearly show that the superior simulator performance of the 5-DIAL and 10-LED groups cannot be attributed solely to the type of heading indicator used nor to the difficulty level of the training criterion alone. Instead, their performance was the result of the unique contribution of type of information display and the training criterion used with it. The concommitant failure to produce significant interactions as a result of the

TABLE 1

Means and Standard Deviations as a Function of Memory Load Level[a]

Measures		Memory Load Levels				
	SD	Control	Minimal	Low	Moderate	High
Heading Errors	.11	3.63[c]	4.05	4.17	4.44	4.38
Overcorrections	1.60	45.72	41.41[d]	34.01	36.01	29.84
Airspeed Errors	.17	4.10	4.36	4.73	4.70	5.07
Rate Errors	3.74	105.23	107.12	105.42	110.35	100.89
Error Time	1.55	6.61	9.41	10.32	14.04	11.83
Memory Scores	1.33	-[b]	95.58	91.14	89.23	83.48

a. All tests were at the .01 level except Error Time (.05 level).
b. There was no score in the control condition.
c. Single underline indicates significance between a control and other levels.
d. Double underline indicates significant differences between memory levels.

TABLE 2

Summary Results on Three Dependent Measures for the Interaction Between Type of Indicator and Level of Training Criterion

	F	P	Omega^2
Heading Errors	7.06 (1,24)	.05	19%
Overcorrections	11.65 (1,24)	.01	21%
Rate Error	7.08 (1,24)	.05	17%
Error Time	8.26 (1,15)	.05	27%

Average Variance Accounted for = 21%

demands of the memory load tasks seems to once again point out the adaptability of the pilot to changing workloads. However, this adaptability did not generalize to the primary task as well for the two other experimental groups, indicating real differences in display type by criterion combinations. Also, contrary to our expectations and previous results, the greatest memory load proved not to be the most decremental on primary task performance. The expected linear function was not found. Presumably, once subjects have sufficient memory load they engage in chunking strategies. The memory load is reduced to fewer items, allowing more residual attention for the primary task.

These interpretations are strengthened by the fact that five of the six dependent measures are convergently consistent. For example, both the 5-DIAL and 10-LED groups produced the least number of heading errors, and, as would be expected, the greatest number of heading overcorrections, indicating these two groups were not only the most accurate in maintaining heading but they were also the most responsive in their attempts to be accurate. This view is further supported by the fact both groups also produced the lowest error time scores beyond the 5 and 10 degree training criterion thresholds. Furthermore, they also produced the most accurate performance in ascent/descent rate, a performance measure strongly related to the heading correction task because, in attempting to maintain attention to heading, the pilot may sacrifice accuracy in his vertical climb or descent.

The relatively high levels of variance accounted for in these primary task interactions suggest strongly the practical significance of relating training criteria to simulator research and equipment design early in the design and development stages. It is quite possible that the frequent and frustrating occurrence of inconsistent results found in some areas of simulator research, for example, in the area of motion effects (see Puig, Harris, and Ricard, 1978) are partially the result of uncontrolled training criterion differences.

References

Bermudez, J. M., Schwank, J.C.H., Longridge, T.M., Smith, B.A., and McCloy, T.M. Effect of peripherally presented visual signals on pilot performance during flight simulation. Wright-Patterson Air Force Base, Ohio; Aerospace Medical Research Laboratory, Tech. Rep. 78-120, June 1978.

Cofer, C.N. Verbal learning. In M.H. Marx (Ed.), Learning Processes. Toronto, Ontario: The MacMillan Company, 1969, 312-313.

Puig, J.A., Harris, W.T., and Ricard, G.L. Motion in flight simulation: An annotated bibliography. Orlando, Florida: Naval Training Equipment Center, Technical Report IH-298, July 1978.

Schwank, J.C.H., Bermudez, J.M., Smith, B.A., and Har s, D.A. Pilot performance during flight simulation with peripherally presented visual signals. Proceedings of the Human Factors Society, 1978, 222-226.

Weitz, J. Criteria for Criteria. American Psychologist, 1961, 16. 228-231.

AUTHOR INDEX

Adams, Jerome
Albornoz, Jaime I.
Allnutt, Martin F.
Anderson, Robert P.
Arima, James K.
Ashworth, B. R.
Barrera, Adriana D
Benel, Russel A.
Bentham, Jack E.
Beringer, Dennis B.
Bermudez, John
Berry, Gene A
Bessemer, David W.
Bickley, William R.
Bittner, Alvah C., Jr.
Bloom, Richard W.
Bloomfield, John R.
Boren, Glenn E.
Brock, J. F.
Bryan, Otis
Calhoon, Janice
Cannon, Ross A.
Carrothers, Scott W.
Castore, Carl H.
Christ, Richard E.
Cicchinelli, Louis F.
Ciley, Colin D.
Clark, Kenneth E.
Clover, William H.
Collyer, Stanley C.
Cornelius, Edwin T.
Corso, Gregory M.
Cowan, Douglas K.
Crisler. Larry J.
Crooks, W. H.
Cummings, William H.
Curtis, Ervin W.
Dalton, Bret
Dansby, Mickey R.
Decker, Phillip J.
DeFleur, Lois
Deignan, Gerard M.
Delligatti, Andrea
Dieterly, Duncan L.
Drummond, Robert J.
Earles, James A.
Frank, Rex A.
Gawron, Valerie J.
George, Clay E.
Glosser, Earl A.

Greenfield, Gary R.
Guerrieri, John A.
Hagman, Joseph D.
Hall, Francine
Hamilton, Kenneth L.
Harburg, Fred D.
Hardwick, Douglas A.
Harris, Cecil B.
Harris, D. A.
Hautman, George M.
Hayes, Fulton D.
Hendrix, Hal W.
Hendrix, William H.
Hooke, Lydia R.
Horowitz
Huelf, Nita
Hughes, Richard L.
Hughes, Ronald G.
Ideen, Dana R.
Jackson, Linda D.
James, Lawrence R.
Jenkins, James L.
Jones, Marshall B.
Jorgensen, Charles C.
Josefowitz, A. J.
Kantor, Jeffrey E.
Kaplan, J. D.
Kaul, Charles E.
Kellogg, Robert S.
Kennedy, Robert S.
Kimball
Kochevar, J. W.
Koonce, Jefferson M.
Krause, Michele
Kuperman, Gilbert G.
Lagarde, Henry J.
Landrum, Cecile S.
Laughery, K. Ronald, Jr.
Lebow, Richard N.
Lintern, Gavan
Luchsinger, V. P.
Mabry, T. R.
Mangelsdorff, A. David
Mann, Dale P.
McCauley, Michael E.
McCloy, Thomas M.
McCord, Douglas
McIntire, Walter G.
Medlin, Steven M.
Mickley, G. Andrew

Miller, Gary G.
Morey, John C.
Mountford, Joy
Nataupsky, Mark
Nita, Patricia C.
O'Connor, Thomas J.
O'Hearn, Thomas P.
Osborne, Arthur D.
Parisher, Darrel
Parker, Kermit E. Jr.
Parrish, R. V.
Pepper, Ross L.
Percival, Lynn C.
Percy, Richard L.
Pfeifer, Jerry R.
Porter, David B.
Potter, Earl H.
Prather, Dirk C.
Prince, Howard T. II
Quartano, Anthony G.
Ricard, G. L.
Richter, Judith
Rinalducci, Edward J.
Roscoe, Stanley N.
Sandry, Diane
Schreiner, William
Schwank, Jock C. H.
Seager
Sells, S. B.
Shepperson, Vance
Short, Lawrence O.
Silsby, Harry D.
Silverman, Lawrence F.
Smith, Deborah A.
Smith, H. Wayne
Smith, Seward
Solomon, Victor M.
Spicher, Charles R.
Stein, Earl S.
Steinheiser, Frederick H.
Stephenson, Matilde K.
Stephenson, Robert W.
Stewart, William C. A.
Stokes, John H.
Sweney, Arthur B.
Sweney, V. Ann
Thompson, Laura A.
Thompson, Paul
Thompson, Thomas J.
Tirman, Valentin

Trehan, Donald E.
Vitola, Bart M.
Wald, Jerry
Warren, James
Watson, Thomas W.
Wells, M. D.
Wickens, Christopher D.
Wiker, Steven F.
Wilcove, Jerry L.
Wilkerson, David A.
Wilson, Lawrence F.
Wood, Frank R.

★ U. S. GOVERNMENT PRINTING OFFICE 1980 - 679-461/6 Reg. 3