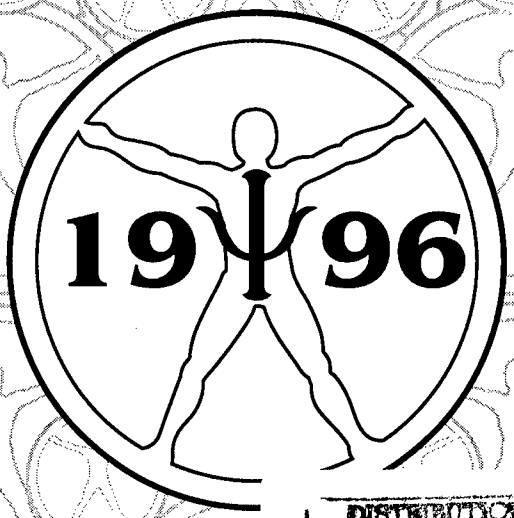
PROCEEDINGS

APPLIED BEHAVIORAL SCIENCES SYMPOSIUM

FIFTEENTH SYMPOSIUM
10-11 APRIL 1996



DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

SPONSOR
USAF ACADEMY DEPARTMENT OF BEHAVIORAL SCIENCES AND LEADERSHIP

19970106 004

DTIC QUALITY INSPECTED 1

PROCEEDINGS OF THE FIFTEENTH BIENNIAL APPLIED BEHAVIORAL SCIENCE SYMPOSIUM 10-11 APRIL 1996

DEPARTMENT OF BEHAVIORAL SCIENCES AND LEADERSHIP
UNITED STATES AIR FORCE ACADEMY
COLORADO SPRINGS, COLORADO

SYMPOSIUM CHAIRS
LT COL JUSTIN D. RUEB
LT COL JOHN MICALIZZI

Applied Behavioral Sciences Symposium

PROFESSOR AND HEAD, DEPARTMENT OF BEHAVIORAL SCIENCES AND LEADERSHIP
Colonel David B. Porter

EDITORS
Lt Col Justin D. Rueb
Captain John D. Garvin

PROGRAM COMMITTEE
Captain John D. Garvin, Chair
Clinical: Major Jeff Jackson
Educational: Capt John D. Garvin
Human Factors: Major Anthony Aretz
Industrial/Organizational: Dr. Ann Herd
Social/Experimental: Dr. Jeanne Smith

OPERATIONS
Captain Derek Abel, Chair
Captain Ronald Merryman
Captain Richard Thul

PUBLICITY
Captain Jeff Nelson
Captain Earl Nason

PROTOCOL
Captain Richard Thul
Captain Lisa Boyce

REGISTRATION
Major Thomas Mabry

LODGING
Major Kenneth Komyathy
Gail Rosado

TRANSPORTATION
Captain Stu Turner
Captain Kirk Broussard

FINANCE
Captain Wesley Olson

CORRESPONDENCE/DATABASE
Captain Terence Andre

Department of Behavioral Sciences and Leadership
USAF Academy, CO 80840

This research report entitled "Proceedings of the Fifteenth Biennial Applied Behavioral Sciences Symposium" 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.

Donald J. McGillen

DONALD J. MCGILLEN
Director of Research

19 Mar 96
Dated

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	10 Apr 96	

4. TITLE AND SUBTITLE
Proceedings, Fifteenth Biennial Applied Behavioral Sciences Symposium

5. FUNDING NUMBERS

6. AUTHOR(S)
Justin D. Rueb, Lt Col, USAF
John D. Garvin, Capt, USAF

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Department of Behavioral Sciences and Leadership
US Air Force Academy/DFBL, 2354 Fairchild Dr., Suite 6L47
USAF academy CO 80840-6228

8. PERFORMING ORGANIZATION REPORT NUMBER
USAFA-TR-96-2

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Unclassified

12b. DISTRIBUTION CODE
Limited

13. ABSTRACT *(Maximum 200 words)*
These proceedings include papers and presentations that deal with a wide range of research in psychology with emphasis on military issues.

14. SUBJECT TERMS
Aircrew and Aviation Issues
Aviation (General)
Cognitive Issues
Equal Opportunity
Educational Assessment
Human Factors Issues

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

FOREWORD

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, Marines, Coast Guard, or of any government organization in which the author(s) or presenter(s) 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 endorsement or approval of the use of such commercial hardware or software.

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

In publishing the Proceedings of the Fifteenth Symposium, the editor sought to facilitate an effective and timely dissemination of the technical information presented. In most cases, the papers contained in this document were printed from electronic copy (edited for format) submitted by the authors, who are solely responsible for their contents.

Acknowledgments

Cover

Graphics Division of Audiovisual Services
Dean of Faculty, USAF Academy

Administrative Support
Gail Rosado
Laura Neal
Marietta Goodman

PROCEEDINGS OF THE FIFTEENTH BIENNIAL APPLIED BEHAVIORAL SCIENCE SYMPOSIUM

FEATURED SPEAKERS LIST xviii

CLINICAL PSYCHOLOGY ISSUES

Assessment of Psychological Factors in Female and Male United States Air Force Pilots 1

Raymond E. King
Suzanne E McGlohn
Paul D. Retzlaff

Exercise as a Protection Against Negative Emotional States 6

Maryanne Martin

Differential Impairment of Naming Latencies for Stress-Related Work 11

Gregory V. Jones
Maryanne Martin

A Profile of a Heavy/Problematic Collegiate Drinker: A Literature Review 17

Michael V. Waggle

Military Psychologist: What is Military Psychology 23

Martin F. Wiskoff
Dana H. Lindsley

Are We Winning the War on Drugs? 28

Clark Hosmer

Using Existential and Cognitive Behavioral Techniques in the Design of a Short-Term Therapy Group for Incest Survivors 32

Timothy P. Kopania

Group Therapy for Rape Survivors: A Combination of Person-Centered and Cognitive-Behavioral Techniques 38

Michael V. Waggle

EDUCATION AND ASSESSMENT ISSUES

Correlates of Course and Faculty Perceived Effectiveness 44

Timothy P. Kopania

Cadet Attitudes About Group Work 52

Justin A. Hansen
David B. Porter

Indicators of Reflective Thinking in College Faculty 58

Kate Preston
David B. Porter

Focus Group Technique as a Classroom Learning Activity 64

Bernard Asiu

Journal Writing: Its Effects on Objective Test Performance in an Upper Division Leadership Class 70

Robert C. Berger
Gary A. Packard, Jr.
Craig A. Croxton

No Pain, No Gain: The Effect of an Intelligent Tutoring System on F-15 Troubleshooting Performance 76

Bradley S. Boyer
Ellen P. Hall
Anna L. Rowe
Robert A. Pokorny

Fostering Students' Motivation in the College Classroom: The Role of Critical Professor Behaviors 82

Ann M. Herd
Lisa Boyce
Randy Stiles
Charlie Law

Two Internal Yardsticks for Integrity 87

Clark Hosmer

Gender and Scholastic Aptitude Test Scores: Relationship to Grade Point Average at the United States Air Force Academy 91

Dawn L. McCown
Justin D. Rueb

Psychological Androgyny and its Relationship to Leadership Grades of Cadets at the United States Military Academy 97

Lori A. Stokan
Sehchang Hah

Panel: An Innovative Approach to Curriculum Evaluation in a Civil Engineering Domain Panel Session 103

Theodore A. Lamb
Winston R. Bennett, Jr.
Kent L. Gustafson
Kurt C. Kraiger
Mike Rits

HUMAN FACTORS ISSUES

Development of an Electronic Cockpit Map Display for Aircraft Ground Navigation 105

Anthony D. Andre
David S. Tu

Airsickness During Flight Training 110

Thomas G. Dobie
James G. May

An Operational Definition and Measurement Method for Situation Awareness 118

Bruce P. Huhn

Subjective Workload Measures: National Aeronautics and Space Administration Task-Load Index in a Task-Saturated Cockpit Environment 123

Keith R. Ober
Anthony J. Aretz

Responses of General Aviation Pilots to Autopilot and Pitch Trim Malfunctions 130

Dennis B. Beringer

The Opto-Kinetic Cervical Reflex (OKCR) in Pilots on High-Performance Aircraft 137

Ronald F. K. Merryman
Anthony J. Cacioppo

Single Seat Fighter Pilot Landing Performance During Multiple, Long-Duration Missions 143

Patrick E. Poole
Daniel H. Bauer
Kory G. Cornum

Hemispheric Dominance and Flight Performance 149

Christopher T. Johannsen
Anthony J. Aretz

Virtual Reality Features of Frame of reference and Display Dimensionality with Stereopsis: Their Effects in Scientific Visualization 156

Edward P. McCormick
Christopher D. Wickens

Spatial Knowledge Acquisition in Virtual Environments 161

Michael J. Singer
Robert C. Allen

Deriving Training Lessons Learned from an Advanced Warfighting Experiment 165

Gary S. Elliott

A Strategy for Efficient Device-Based Tank Gunnery Training in the Army National Guard 170

Joseph D. Hagman

Training on Simulators and Live Fire Platoon Gunnery Performance 175

Bruce Sterling

Panel: Psychomotor Abilities 179

Patrick C. Kyllonen
Scott R. Chaiken
Joshua B. Hurwitz
Lee Gugerty

LEADERSHIP AND ORGANIZATIONAL DEVELOPMENT ISSUES

Behavioral Indicators of Effective Performance and Leadership as Identified Through a Policy-Capturing Method 181

Linda S. Hurry
Guy S. Shane
James R. Van Scotter

Cognitive Therapies for Intelligent Organizations 187

John R. Landry

Reengineering the Human Interface with Space: A Team Approach to Process Improvement 193

Frank McIntire
Chip Houlihan

Reengineering Our Organizations: A Leadership Challenge 200

John Micalizzi

Panel: Executive Coaching: How to Achieve Long-Term Leadership Behavior Change 206

Gordon J. Curphy
Vonda K. Mills

Panel: Bridging the Gap Between Leadership Outcome Development and Their Assessment 207

R. R. Albright
P. T Kelly

Panel: The Fundamental Role of Leadership: Developing Followers into Partners 211

William E Rosenbach
Thane S. Pittman
Earl H. Potter III

Panel: New Horizons in Leadership: Creating Organizational Rainbows 212

Imagination 213

Frank Prochaska

Creativity and Innovation in Change Leadership 214

Vicki L. Strunk

Creativity: The New Bottom Line 215

Bill Wallisch

Re-creation, not Wreck-reation 216

Don Marble

Vaulting Barriers to Creativity 217

Jerry Reinsma

Self-Discovery 218

Betty Rosengren

Panel: Personality and Leadership: What Do We Know About Selection, Training, and Development 219

Gordon J. Curphy
Kevin D. Osten
Jeffrey Voetberg

The Leadership Development Survey in a Reserve Officer Training Corps Setting 220

Jeffrey Voetberg

INDUSTRIAL PSYCHOLOGY ISSUES

Effects of Proximal and Distal Context Variables on Performance Appraisal Quality: A Model and Framework for Research 227

Kevin R. Murphy
Jeanette N. Cleveland
Christine Henle
Kim Morgan
Michael Orth

Evaluation of a Reengineered Performance Appraisal and Reward System Within the Federal Government 235

Steven R. Frieman

The Relation of Prior Performance Feedback Ratings to Managers' Subsequent Feedback Seeking Behavior 240

Ann M. Herd

Factors Contributing to the Morale, Cohesion, and Motivation of Combat Support Personnel During Desert Shield/Desert Storm 246

Gary Jandzinski
David Vaughan
Jim Van Scotter

Evidence of the Usefulness of the Trait of Agreeableness for Selecting Employees to Reduce Performance Variability in Critical Group Tasks 251

Max R. Massey
James R. Van Scotter
Guy R. Shane

Combat and Non-Combat: Should Individual Values Differ? 256

Herbert G. Baker

The Relationship Between Environmental Attitudes and Environmental Behaviors Among Air Force Combat Command Members 261

Daniel T. Holt
Steven T. Lofgren
Guy Shane
Kevin L. Lawson

Estimating the Utility of Organizational Change Using Probability-Based Simulations 268

Winston Bennett, Jr.
Robert M. Yadrick
Bruce Perrin

Computer Adaptation of Task-Based Occupational Analysis to the Changing World of Work 273

William J. Phalen
Jimmy L. Mitchell

Task-Based Analysis of Processes 279

Brice M. Stone
Kathryn L. Turner
Robert C. Rue
Sharilyn A. Thoreson
Jmmy L. Mitchell

Analysis of Outcomes for an Entity Based Job and Training Simulation Model 284

Kathryn L. Turner
Brice Stone
Guy Curry
Teresa Bennett

Air Force Occupational Measurement Squadron Customer Satisfaction Survey Report: A Summary of Findings Regarding Examinee Knowledge of the Testing Portion of the Weighted Airman Promotion System 289

Heather M. Henderleiter

Quality of Life in the United States Army Recruiting Command 293

H. Michael Hughes

Ability of Military Recruits: 1950 To 1994 and Beyond 298

Brian K. Waters
Dana H. Lindsley

Quality of Life in the Navy 305

Gerry L. Wilcove
J. Philip Craiger
Joyce S. Dutcher

The Senior Leader Equal Opportunity Survey: What Do the Bosses Believe? 310

Mickey R. Dansby

The Effects Of Race on Procedural Justice: The Case of the Uniform Code of Military Justice 317

Dan Landis
Michael Hoyle
Mickey R. Dansby

The Relationship Between Racism/Sexism and Group Cohesiveness and Performance 322

Robert E. Niebuhr
Stephen B. Knouse
Mickey R. Dansby
Katherine E. Neibuhr

Harassment in the Canadian Air Force: 1992 and 1995 Survey Results 328

Brian R. Thompson

Hierarchical Classification of Training Needs Affecting Flight Crew Performance 334

Lawrence L. Bailey
Rogers V. Shaw

Towards a Unified Theory of Airmanship: A Model for Education 340

Tony Kern
J. D. Garvin

An Evaluation of Full Flight Simulators and Flight Training Devices in Air Carrier Initial Flight Training Programs 346

John Wolf
Gerald Gibb
Steven Hampton
John A. Wise

Shaping Tomorrow's Military: The National Agenda and Youth Attitudes 353

W. S. Sellman
William J. Carr
Dana H. Lindsley

Panel: Determinants of Military Allied Health Care Students Success: A Multifactorial Analysis 356

Russell D. Porter
Jimmy L. Sterling
Joy P. Vroonland
Squy G. Wallace

Panel: Life Aboard a U. S. Aircraft Carrier: Examinations of Biomedical and Safety Issues 357

Robert Stanny

The Stress and Strain Associated with Deployment Aboard a U. S. Aircraft Carrier 358

Douglas A. Wiegman

Work/Rest Cycles and Strain Among Flight Deck Personnel Aboard a U. S. Aircraft 359

David McKay
Dylan Schmorrow

The Naval Flight Deck: An Unforgiving Environment for the Untrained or Complacent 360

Scott Shappell

Quality of Life and Perceptions of Stress and Strain: An Examination of Flight Deck Crew Interviews 361

Dylan Schmorrow
Claire Portman
David McKay

Panel: Team Effectiveness in the Space Launch Environment: Theory to Application 362

Jeffrey S. Austin
Robert C. Ginnett
Barbara G. Kanki
Cheryl M. Irwin
Earl R. Nason
Timothy S. Barth
Patrick S. Simpkins
Donna M. Blankmann-Alexander
Mark J. Nappi

Panel: Recent Developments in Methods for Formative Evaluation in Military Education 363

Winston R. Bennett, Jr.
Kent L. Gustafson
William Wheeler

Panel: Predicting Crew Resource Management (CRM) Aspects of Aircraft Commander Performance Using a Situational Judgment Test 364

Kenneth T. Bruskiewicz
Jerry W. Hedge
Mary Ann Hanson
Kristi K. Logan
Walter C. Borman
Frederick M. Siem

Panel: Behavioral Sciences Career Field Review 365

William H. Cummings, III

POSTERS

Educational Innovations: Advanced Technology Assessment 366

Todd A. Fore

A Leadership and Communication Skills Development Training Program for Airport Checkpoint Security Supervisors: Development and Evaluation 367

Gerald D. Gibb
Sam C. Kelly
James S. Baker
Daniel Sola
Colleen Wabiszewski
Xavier Simon

Training Ship Handling Skills in a Virtual Environment: A Comprehensive Requirements Determination Approach 368

Robert T. Hayes
Rosemary Garris-Reif

Effects of Fitness and Individual Characteristics on Cardiovascular Reactivity and CHD Potential 369

William H. Hendrix
Richard L. Hughes

Content Analysis of Employees' Reports of Their Feedback Seeking Behavior 370

Ann M. Herd
Heather Pringle

Low-Visibility Surface Operations: Crew Navigation Strategies and Use of Taxi Maps 371

Cheryl M. Irwin
Kim E. Walter

Using Gender Role Conflict Scores to Predict Requests for Different Types of Counseling Services 372

R. Jeffrey Jackson
Christopher R. Kieling
Jonathon R. Eckerman

Social Change and the Eating Habits of Air Force Enlisted Personnel 373

Stephen J. Jirka
Edna R. Fiedler
Heather Ktenidis
R. Brian Howe
William G. Jackson
Diane Cortner

Training Perishable Perceptual Skills 374

Steven J. Kass
Robert H. Ahlers

Summative Evaluation of OPS/CEAF FERL 375

Kurt C. Kraiger
Mark Teachout
Theodore A. Lamb

Application of the Critical Incident Technique to Enhance Crew Resource Management Training 376

Kristi K. Logan
Mary Ann Hanson
Jerry W. Hedge
Kenneth T. Bruskiewicz
Walter C. Borman
Frederick M. Siem

Perceived Job Security and Veteran Status: Effects of Type of Occupation 377

Michael D. Matthews
Charles N. Weaver

Using Pre-Course Reflective Judgment and Critical Thinking Measurement Instruments as Guides for Lesson Preparation 378

Ronald F. K. Merryman
Anthony J. Aretz

Characteristics of Sleep, Mood, and Performance Patterns in Battalion Staff Members at the Joint Readiness Training Center 379

Robert J. Pleban
Tina L. Mason
Patrick J. Valentine

Scoring Work Sample Data of Complex Ill-Structured Tasks with Expert Holistic Judgments 380

Robert A. Pokorny

Implementation of Assessment Measures and Curriculum Integration 381

Michael P. Rits
Mark Teachout
Theodore A. Lamb

Short Term Effects of Acceleration on Human Subjects 382

Dylan Schmorrow
James Siwert
David Moyers

An Enlistment Screen for Non-High School Graduates: Development, Operational Test, and Evaluation 383

Thomas Trent

Application of Sequential Data Analysis Techniques to the Instructional Domain 384

Brenda M. Wenzel

INDEX TO AUTHORS 385

Applied Behavioral Sciences Symposium

Featured Speakers

Keynote Address:

Dr. Peter Salovey
Professor of Psychology
Yale University
"Emotional Intelligence"

Invited Speakers:

Dr. Daniel Ilgen
Professor of Psychology
Michigan State University
"Decision Making in Teams:
Performance and Motivational Issues"

Dr. Robert North
Engineering Psychology, Ph.D.
Honeywell-Minneapolis
"Strategic Planning for 21st Century
Human-Centered Systems Research
and Development: A Control Company's View"

Assessment of Psychological Factors in Female and Male United States Air Force Pilots[1]

Major Raymond E. King, Psy.D.
Major Suzanne E. McGlohn, M.D.
Armstrong Laboratory

Paul D. Retzlaff, Ph.D.
University of Northern Colorado

Abstract

We studied psychological traits found in 64 male and 50 female nonpsychiatrically referred United States Air Force pilots. Participants completed computerized versions of the Multidimensional Aptitude Battery (MAB) and the NEO Five Factor Inventory (NEO-FFI) and were administered a semi-structured interview. While MAB IQs were near identical for men and women, women were found to have higher scores on Extraversion, Agreeableness, and Conscientiousness on the NEO-FFI. The semi-structured interview suggested the United States Air Force Academy is an important avenue for women to enter military aviation. An important potential training issue may be men's desire to protect women in combat.

Although female aviators have been an integral part of military aviation since World War II, little is known scientifically about their psychological make-up. Women comprise a small, but growing, percentage of United States Air Force (USAF) pilots (2% or approximately 315 as of Jan 95). Novello and Youssef (1974) studied 87 general aviation female pilots and found female pilots to be more similar to male pilots than to females in the general population. Jones recognized a need to study female aviators as early as 1983, publishing an alert to flight surgeons on the stress of the conflicting roles that female aviators face. Due to the decision to open up almost all USAF jobs to women, identification of the stresses of mixed-gender squadrons, attention to the psychological concerns of pilots in combat, and recognition of the difficulties of balancing a career and family are important in today's USAF.

---

[1] This work was supported by the U.S. Army Medical Research and Materiel Command.

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

The investigators adhered to policies regarding the protection of human subjects as prescribed by 32 CFR 219 and Subparts B, C, and D.

The structure of the paradigm of the "Right Stuff" (Retzlaff & Gibertini, 1987; Siem & Sawin, 1990; Wolfe, 1980) rests on a male foundation. Do female pilots, however, bring different intellectual skills and personality styles into the cockpit? Siem and Murray (1994) found that experienced pilots rated "conscientiousness" as the most important of the "big five" (neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness) personality characteristics determining pilot performance. Siem and Murray advocate research to validate the importance of conscientiousness in actual pilot performance. Also, scores from intelligence testing could establish a range of cognitive capabilities in successful female aircrew, leading to improved screening of female pilot candidates.

Method

Participants

One hundred and fourteen pilots (64 men and 50 women) from Air Mobility Command (AMC) and Air Education and Training Command (AETC) participated in the present study. Most participants (n= 108) were assigned to crewed aircraft (transport/tanker; C-5, C-17, C-21, C-141, KC-10, KC-135). All AETC (n = 6) participants were instructor pilots who had a recent history of assignment to crewed aircraft.

Table 1.

Demographics

	Women*	Men
Mean age	30.25	29.33
Mean self-reported military flying hours	1,760.00	1,712.11
Mean self-reported combat-support flying hours	43.20	67.83
Race**	(Expressed	as percents)
Asian	0	1.60
Black	2.04	6.25
Caucasian	97.96	89.06
Other/Wouldn't Identify	0	1.6
Married		
Yes	53.10	67.19
No	46.90	32.81
Education		
Bachelors	44.90	53.13
Some Grad Work	22.45	34.38
Masters	32.65	9.65
More than 18 years	0	3.13

Table 1. Demographics (Continued)

	Women	Men
Commissioning source	(Expressed	as percents)
OTS	12.24	15.63
ROTC	30.61	45.31
USAFA	55.10	39.06
MIMSO	2.04	0
Military Rank		
O-2	12.24	9.38
O-3	71.43	87.50
O-4	6.12	3.13
O-5	8.16	0
O-6	2.04	0
Crew position		
Co-pilot	40.82	31.25
Pilot	20.41	42.19
Aircraft Commander	16.33	9.38
Instructor Pilot	18.37	10.94
Stan Eval	4.08	6.25
Private Pilots' License		
Yes	67.35	65.63
No	32.65	34.38

* Due to disk failure, includes demographic information on 49 of the 50 female participants.

** English first language for all participants (necessary information for the MAB).

Apparatus

We used six IBM ThinkPad dual-scan color notebook computers (486DX with 8 Meg RAM) capable of writing entries onto a 3.5 inch DSHD disc. Computer administration allows confidentiality and anonymity, as well as standardization.

Procedure

We solicited data within flying squadrons from non-psychiatrically referred USAF pilot volunteers. A female psychiatrist conducted a semi-structured clinical interview to provide information about personal health, family health, squadron relationships, and career or deployment

stresses. The interview covered the impact of grounding greater than thirty days, motivation to fly, health decrements due to aircraft design, teamwork difficulties and blocks to success, career demands, combat and prisoner-of-war concerns, stress and coping styles, flying goals, and family and health concerns.

We tested from one to six participants at a time, collecting general demographic information and administering the Multidimensional Aptitude Battery (MAB; Jackson, 1984) and the NEO Five Factor Inventory (NEO-FFI; Costa & McCrae, 1992). The MAB is an intelligence (IQ) test; the NEO-FFI is a survey of the normal range of personality functioning.

Results

Of the ten subtests of the MAB, only Information and Picture Completion showed significant female/male differences (with males higher). Female pilots achieved a Verbal IQ equal to 120.0 (5.4 SD), Performance IQ equal to 121.7 (7.0 SD), Full Scale IQ equal to 122.3 (5.2 SD) while male pilots achieved 120.8 (5.6 SD), 122.7 (7.3 SD), and 123.4 (5.4 SD), respectively. There were no significant male/female IQ differences. Women scored higher on the NEO-FFI domains of Extraversion [$M = 62.44$, 10.11 SD women; $M = 58.06$, 11.04 SD men; $t\ (110) = 2.15$, $p < .05$], Agreeableness [$M = 54.29$, 9.86 SD women; $M = 47.44$, 11.15 SD men; $t\ (110) = 3.38$, $p < .001$], and Conscientiousness [$M = 55.6$, 10.06 SD women; $M = 51.34$, 9.52 SD; $t\ (110) = 2.29$, $p < .05$]. Combined-gender norms, as opposed to separate male and female norms, were used in calculating standard (T) scores to facilitate gender comparisons.

While men (50%) reported on interview that they had wanted to be a pilot since childhood, women (36%) reported that they became interested in flying upon attending the Air Force Academy and learning they were pilot qualified. The majority of men (76.56%) reported frequent absences have strained their relationships, while only 50% of women reported similar relationship strains. Participants reported that the squadron members with the most difficulty dealing with women in the squadron are older males, including enlisted crew, and some commanders. Finally, 73% of men reported they would be more protective of a female in combat.

Discussion

The flying community is atypical of the general population as demonstrated by the high average to superior IQ and small standard deviations due to multiple selection and self-selection forces. Incumbent female pilots seem to have even more of a "good thing" in terms of positive personality traits. Occupational norms for non-referred pilots may be helpful in future pilot selection, assignment, and retention decisions.

While these male participants may have selected their aircraft based on their preference to work as part of a crew (and hence self-selected), most of these female participants did not have many options, due to the combat exclusion law in effect at the time of their assignment. The United States Air Force Academy appears to be an important avenue for women to enter a military aviation career. Men's desire to protect women in combat and in a prisoner of war scenario may be an important training issue.

References

Costa, P. T., & McCrae, R. R. (1992). Professional manual Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI). Odessa, FL: Psychological Assessment Resources, Inc.

Jackson, D. N. (1984) Multidimensional Aptitude Battery manual. London, Ontario, Canada: Research Psychologists Press, Inc.

Jones, D. R. (1983). Psychiatric assessment of female aviators at the U.S. Air Force School of Aerospace Medicine (USAFSAM). Aviation, Space, and Environmental Medicine, 54 (10), 929-931.

Novello, J. R., & Youssef, Z. I. (1974). Psycho-social studies in general aviation: II. Personality profile of female pilots. Aerospace Medicine, 45 (6), 630-633.

Retzlaff, P. D. & Gibertini M. (1987). Air Force pilot personality: Hard data on "The Right Stuff." Multivariate Behavioral Research, 22, 383-399.

Siem, F. M., & Murray, M. W. (1994). Personality factors affecting pilot combat performance: a preliminary investigation. Aviation, Space, and Environmental Medicine, 65 (5, Suppl.), A45-A48.

Siem, F. M., & Sawin, L. L. (1990, April). Comparison of male and female USAF pilot candidates. Paper presented at the 69th Symposium of the Aerospace Medical Panel, Tours, France.

Wolfe, T. (1980). The right stuff. New York: Bantum Books.

Exercise as a Protection against Negative Emotional States

Maryanne Martin, DPhil
University of Oxford, UK

Abstract

Many questions are at present unanswered concerning the relations between exercise and psychological factors. The present study focuses on exercise and emotional state. Exercise is assessed by means of sporting activity, and the psychological factors which are studied include not only a range of emotions but also personality and reasons for taking exercise.

In many people's minds one of the best indicators of being healthy is being physically fit (Blaxter, 1990; Radley, 1994). Studies suggest that regular exercise decreases risk of a number of life-threatening illnesses, including coronary heart disease and cancer (Blair et al., 1989) as well as helping in the management of diabetes, obesity and depression (Koplan et al., 1989). In chronic fatigue syndrome patients feel very tired and the reluctance to take physical exercise is central to the diagnosis of this disorder (Fry & Martin, in press). On the other hand, excessive exercise can be detrimental to health, for example in anorexia nervosa where vigorous exercise may be used to lose weight or among joggers who run more than 50 miles per week and may risk decalcified bones, decreased bone mass, stress fractures and scoliosis (sideways curvature of the spine), Sapolsky (1994). Is it possible to demonstrate empirically that regular exercise provides protection against negative emotional states in healthy young adults? There is some evidence that exercise can indeed act psychologically as a buffer against stressful events (Brown & Lawton, 1986; Brown & Siegel, 1988; Lindsay & Powell, 1994). However, a number of questions remain concerning these effects. How specific are the effects of exercise with respect to different emotions? Are they also dependent upon the personalities of the persons concerned and their reasons for taking exercise? These questions are addressed by the present study.

A major way in which people take regular exercise is via sporting activity. This is particularly so for a student population. Thus an experiment was carried out in which the sporting activities of a group of students were assessed and the relations between such exercise and the participants' emotions and personalities were elucidated.

Method

Subjects

There were 140 subjects of whom 62 were female and 78 were male. They were midway through their first term at the University of Oxford. The mean age of the sample was 19.7 years with a standard deviation of 2.5 years.

The author wishes to thank Christopher Gent and Claire Woolley for help in testing the subjects.

Materials

Mood and personality were assessed using questionnaires. Depression was measured for the preceding week with the Beck Depression Inventory (BDI: Beck et al., 1961). A comparable scale was used for happiness, the Oxford Happiness Inventory (OHI: Brebner & Martin, 1995) developed by the author in collaboration with Argyle and Crossland. Trait anxiety was measured using Spielberger's trait anxiety scale (Spielberger et al., 1983). Personality was measured using Eysenck's Personality Questionnaire (EPQ: Eysenck & Eysenck, 1991).

A sports questionnaire was devised to assess physical exercise. Subjects were instructed to interpret the term "sport" as widely as possible to include, for example, not only football and swimming but also mountaineering, aerobics, jogging, dancing, rambling, and yoga. Each sport participated in was listed by the subject together with the number of hours per week during relevant months spent playing or training for this sport, the number of months per year playing or training for this sport, and finally the level of attainment: recreational only, college (or school) team, university (or county) team, or national team. Reasons for doing sports were assessed by four 100-point scales ranging from 0 (not at all for this reason) to 100 (purely for this reason). The reasons assessed were because "I like to compete", "I like to be fit", "I like to meet other people", and "I like the sense of achievement".

Results

Of the 140 participants, 5 did no sport at all, 32 did sport for recreation only, 56 for inter-college or school teams, 38 for university or county teams and 8 for national teams. Two measures of sport were used, the total number of hours of sport per year and the level of attainment. As was expected, those who had reached a higher level of attainment spent more time on sport, $r(139) = .52$, $p < .001$.

As shown in Table 1, subjects who spent more time on sport tended to be happier and less anxious. Those with higher levels of sports attainment tended to be happier and have less addictive personality types.

Table 1
Relations between Sport Activity, Mood, and Personality

Sport Activity	Happiness	Trait Anxiety	EPQ:Addiction
Hours per year	.22**	-.18*	-.05
Level of attainment	.18*	-.15	-.20*

In all tables, *** is $p < .001$, ** is $p < .01$, * is $p < .05$

As shown in Table 2, subjects who spent longer on sport or had achieved higher levels in sport tended to do so primarily because they liked to compete and because of the sense of achievement.

Table 2
Relations between Sport Activity and Reasons for Doing Sport

Sport Activity	Compete	Fit	People	Achieve
Hours per year	.36***	.09	.17*	.20*
Level of attainment	.46***	.13	.16	.31***

As shown in Table 3, happy and extravert subjects tended to do sport to meet other people. Happier, less anxious, less neurotic, less addictive and less criminal personalities do sport to keep fit.

Table 3
Relations between Mood and Personality and Reasons for Doing Sport

Mood and Personality	Compete	Fit	People	Achieve
Happiness	.12	.21*	.29***	.16
Trait anxiety	-.11	-21*	-.11	.04
EPQ: Extraversion	.08	.17	.30***	.05
EPQ: Neuroticism	-.08	-.18*	.00	.09
EPQ: Addiction	-.12	-.25**	-.12	-.08
EPQ: Criminality	-.07	-.20*	-.07	-.01

Subjects who spent longer on sport tended towards a more positive attributional style when ascribing causes to events where the cause may be positive, negative or neutral, $r(134) = .28$. $p < .001$. They also ruminated less about something which has gone wrong in their lives, $r(137) = -.19$, $p < .05$. Instead they tended to have a more problem-solving approach, $r(137) = .21$, $p < .05$. Of the reasons for doing sport, individuals who did sport to keep fit had a more positive attributional style, $r(130) = .29$, $p < .001$. When something went wrong in people's lives, those who did sport to keep fit, to meet people, or for the sense of achievement made an effort not to think about it, $r(132) = .18$ ($p < .05$), .31 ($p < .001$), and .25 ($p < .01$), respectively. Also those who do sport for the sense of achievement tended to ruminate more, $r(132) = .21$, $p < .05$, but also tended to have a more problem-solving approach, $r(132) = .29$, $p < .001$. Styles of thinking about something that has gone well in subjects' lives were not significantly related to amount of sport or reasons for doing it.

Finally there was no significant effect of gender upon the number of hours of sport per year, level of attainment, mood or personality variables. In multiple regression analyses, when gender was added as a main effect or as an interaction term with another independent variable these factors failed to reach significance.

Discussion

The first term at university is a stressful time for students. For many it is the first time they have lived away from home or alternatively without the imposed life-style discipline of a boarding school. It was found that first-year students who engaged in a relatively large amount of sporting activity tended to be happier and less anxious. When presented with events where the cause was ambiguous they tended towards a positive and away from a negative interpretation. When something had gone wrong in their lives they tended not to dwell on it, but instead to adopt a problem-solving approach.

People's reasons for doing sport were also found to be related to mood and personality. Those who did sport to meet people tended to be happier and have an extravert personality, as well as attempting not to ruminate on negative life events. Those who did sport to keep fit also tended to be happier, less anxious, to have less neurotic, addictive and criminal personalities, and attempted not to ruminate.

It thus appears that engaging in sporting activities at university does indeed help to protect some students from some of the stressful events during the first year. In principle, there could be physiological, social, or cognitive pathway mechanisms for this. Physiologically, it could be that sport leads to an increase in release of endorphins, or affects catecholamine production which is sensitive to changes in stress level. Socially, sport involves meeting other people and this itself is strongly linked to happiness. Finally, the possibility explored in more detail in this study is that cognitive style is different in sporting individuals. Their more positive attributional style, their reduced rumination concerning their own negative life events, and their greater focus on problem solving when something goes wrong may well serve to protect them from negative psychological consequences.

References

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. Archives of General Psychiatry, 4, 561-571.

Blair, S. N., Kohl, H. W., Paffenberger, R. S. Jr, Clark, D. G., Cooper, K. H., & Gibbons, L. W. (1989). Physical fitness and all-cause mortality: a prospective study of healthy men and women. Journal of the American Medical Association, 262, 2395-2401.

Blaxter, M. (1990). Health and lifestyles. London: Tavistock/Routledge.

Brebner, J., & Martin, M. (1995). Testing for stress and happiness: The role of personality factors. In C. D. Spielberger & I. Sarason (Eds.), Stress and emotion (Vol. 15, pp. 139-172). Washington, DC: Taylor & Francis.

Brown, J. D., & Lawton, M. (1986). Stress and well-being: The moderating role of exercise. Journal of Human Stress, 12, 125 -131.

Brown, J. D., & Siegel, J. M. (1988). Exercise as a buffer to life stress. A prospective study of adolescent health. Health Psychology, 7, 341-353.

Fry, A. M., & Martin, M. (in press). Cognitive idiosyncrasies among children with the chronic fatigue syndrome: Anomalies in self-reported activity levels. Journal of Psychosomatic Research.

Eysenck, H. J. & Eysenck, S. B. G. (1991). Manual of the Eysenck personality scales. London: Hodder & Stoughton.

Koplan, J. P., Caspersen, C. J., & Powell, K. E. (1989). Physical activity, physical fitness, and health: Time to act. Journal of the American Medical Association, 262, 2437.

Lindsay, S. J. E., & Powell, G. E. (1994). The handbook of clinical adult psychology. London: Routledge.

Radley, A. (1994). Making sense of illness: The social psychology of health and disease. London: Sage.

Sapolsky, R. M. (1994). Why zebras don't get ulcers: A guide to stress, stress-related diseases, and coping. New York: Freeman.

Spielberger, C. D., Gorsuch, R. L. Vagg, P. R., & Jacobs, G. A. (1983). Manual for the state-trait anxiety inventory. Palo Alto, California: Consulting Psychologists Press.

Differential Impairment of Naming Latencies for Stress-related Words

Gregory V. Jones, Ph.D.
University of Warwick, UK
Maryanne Martin, D.Phil.
University of Oxford, UK

Abstract

Are different levels of psychological stress associated with particular patterns of cognitive processing? For each member of a group of industrial managers, stress was assessed by means of the Stress Arousal Checklist and cognitive performance was assessed by means of a modified Stroop task. The Stroop task employed neutral words or negative, stress-related words such as "deadline". It was found that color-naming latencies for negative words, unlike those for neutral words, were specifically impaired for high-stress individuals, compared to low-stress individuals. The implications of this link between stress and cognitive processing are briefly explored.

Considerable theoretical controversy surrounds the concept of stress, and in particular its psychological measurement (e.g., Pearlstone, Russell, & Wells, 1994). Nevertheless, it has frequently been argued that the level of stress a person suffers influences their state of well-being. People have to contend with many types of difficulty in their everyday existence. If not dealt with successfully, these different problems may tend to induce a common set of changes in people's physical and mental constitutions. People experiencing difficulties in this way can be described as suffering from high levels of stress. A major source of such difficulties resides for many people in aspects of their employment, and thus there has been considerable interest in the possibility that occupational stress is an important determinant of a person's well-being both in the general population (e.g., Arsenault & Dolan, 1983; Cooper & Marshall, 1976; Fisher, 1986) and among specific groups such as dentists (e.g., Cooper, Watts, Baglioni, & Kelly, 1988; DiMatteo, Shugars, & Hays, 1993).

How should levels of stress be assessed? One approach which has been successfully adopted is that of developing self-report measures (e.g., Cooper, Sloan, & Williams, 1988; Cox & Mackay, 1985; Mackay, Cox, Burrows, & Lazzerini, 1978). These measures rely upon an individual's conscious awareness and veridical reporting of stress-related factors. An important alternative to consider is that the effects of stress may be assessed in terms of their behavioural concomitants. An assessment instrument of this type would rely upon objective performance rather than upon self report. In such a case, the significance and interpretation of response patterns is unlikely to be transparent and therefore possible demand effects should be minimised.

The authors thank Emily Woodfield for her contribution to the work reported here.

A good candidate for use as a behavioral index of stress is the Stroop task. In the Stroop task (see MacLeod, 1991), it is found that the latency with which an ink colour can be named increases when the ink forms a conflicting color name (e.g., when red ink forms the word "blue"). In a variation of this task, a reduced effect may be obtained using words other than color-names. In particular, the existence of emotion-specific Stroop interference has been demonstrated (seeWilliams, Watts, MacLeod, & Mathews, 1988). For example, the average latency to name the ink-color of threat-related words tends to be greater for a person with high anxiety than for one with low anxiety (Martin, Williams, & Clark, 1991). Similarly, someone who has a fear of spiders will tend to have higher latencies for naming the ink-color of words such as "cobweb" (Martin, Horder, & Jones, 1992; Martin & Jones, 1995; Watts, McKenna, Sharrock, & Trezise, 1986).

Given the success of the modified Stroop task in probing the interaction between emotional state and cognitive performance, it may also provide a window through which to examine people's levels of stress. That is, it is possible that with the use of appropriate stimuli a modified Stroop task might provide an objective index of individuals' levels of stress.

Method

Subjects

The subjects were 40 managers (38 male, 2 female) employed by a large UK industrial company and had an average age of 40 years.

Apparatus

Materials were constructed for use in a modified Stroop task. There were three types of letter-string stimuli. Negative stimuli consisted of nouns relating to stress, selected from an article by Cooper and Marshall (1976): DEADLINE, EXHAUSTION, FAILURE, OVERLOAD, and REDUNDANCY. Neutral stimuli consisted of nouns matching in word frequency (Kucera & Francis, 1967), number of syllables, and number of letters: FAIRNESS, ADMITTANCE, BALANCE, SANCTITY, and ESTIMATION. Control stimuli consisted of strings of the letter O, again matched on numbers of letters (i.e., OOOOOOOO, etc.). Stimuli were printed in five different colors: red, blue, green, orange and brown. A card was prepared for each condition, as used by Williams and Broadbent (1986). Each card had two columns of 25 stimuli each. The 50 stimuli for each condition contained 10 instances of each of the relevant stimuli and 10 instances of each color, in a randomised order.

Individual levels of stress were assessed using the Stress Arousal Checklist (SACL). The SACL (Cox & Mackay, 1985; Mackay et al., 1978) elicits ratings of current feelings on a four-point scale with respect to 30 adjectives such as "bothered" (positive stress loading) and "peaceful" (negative stress loading.

Procedure

Subjects were tested individually, with administration of the modified Stroop task following that of the SACL. Subjects were instructed to name the colours of the 50 Stroop items in each condition as quickly as possible, and timed with a stopwatch. Latin squares were used to balance the order of the three conditions across subjects.

Results

Subjects were divided using a stem-and-leaf procedure (Tukey, 1977) into High-stress and Low-stress groups. High-stress subjects has an SACL score in the range 10-16 ($N = 9$, mean = 12.3, $SD = 1.9$) whereas low-stress subjects had a score in the range 0-8 ($N = 31$, mean = 3.7, $SD = 2.8$).

Mean Stroop latency data are shown in the first three data columns of Table 1. Analysis of variance with Group (High-stress, Low-stress) and Condition (Negative, Neutral, Control) as between-subjects and within-subject factors, respectively, yielded no significant effect of Group, $F(1,38) = 1.49$, but a significant main effect of Condition, $F(2,76) = 37.13$, $p < 0.001$. Most noteworthy was that the interaction between Group and Condition was also significant, $F(2,76) = 3.41$, $p < 0.05$.

Table 1

Mean Latencies and Latency Differences (sec) for Low-stress and High-stress Groups

	Latency			Latency difference	
Group	Negative	Neutral	Control	(Negative - Control)	(Neutral - Control)
Low-stress	37.84	37.91	33.26	4.58	4.65
High-stress	43.18	41.11	34.16	9.02	6.95

The significant interaction was investigated via a Bonferroni analysis. Latency difference scores were calculated (see final two columns of Table 1), and it was found that for Negative stimuli this Stroop interference effect was significantly greater for the High-stress than the Low-stress group, $F(1,38) = 5.57$, $p < 0.05$, whereas for the Neutral stimuli there was no significant difference between the two groups, $F(1,38) = 2.27$.

Regression analysis showed the same pattern of results. For Neutral stimuli the latency difference scores were not significantly dependent upon SACL scores, $F(1,38) = 3.00$. For

Negative stimuli there was in contrast a significant relation, $F(1,38) = 6.62$, $p < 0.05$; the best-fitting equation was $L = 0.454S + 3.00$, where L is latency difference and S is SACL score.

Discussion

The experiment reported here demonstrates a link between objective patterns of performance in a cognitive task and levels of stress assessed via subjective report. High-stress individuals were specifically impaired in naming stimulus color in a modified Stroop task when the stimulus comprised a stress-related word such as "deadline" rather than a neutral word such as "fairness". Previous work on "cognitive biases" has generally concentrated on their relation to emotional state (e.g., Martin et al., 1991) rather than to level of stress (though see Mogg, Bradley, & Hallowell, 1994). The present finding raises a number of interesting possibilities.

First, the observation of a link between stress and patterns of cognitive processing suggests that cognitive intervention may be an appropriate avenue by which to seek to control stress. The fact that the present study was carried out in an industrial workplace setting lends some support to the possibility that it may be fruitful to approach occupational stress factors in this way. A cognitive therapy for stress can be envisioned, analogous for example to that devised for panic attack (see Clark, Salkovskis, Gelder, Koehler, Martin, Anastasiades, Hackmann, Middleton, & Jeavons, 1988).

Second, the cognitive task employed here may be of service as an object index of stress. One context in which such a measure might be useful is in monitoring the effectiveness of procedures aimed at reducing stress (e.g., Ivancevich, Matteson, Freedman, & Phillips, 1990; Murphy, 1984; Reynolds, Taylor, & Shapiro, 1993), in the same way that another version of the Stroop task has been employed to monitor the course of desensitisation treatment for phobia (Watts et al., 1986).

How secure are the conclusions drawn from the present study? Clearly, there are a range of experimental variables which can be manipulated to investigate their robustness. However, it is helpful that their interpretation is not dependent upon any specific theoretical interpretation of performance on the Stroop task, for example that of Glaser and Glaser (De Houwer & d'Ydewalle, 1994; Glaser & Glaser, 1989). Similarly, presentation format (card-based versus computer-based) has been suggested as an important factor in the modified Stroop task (Dalgleish, 1995), but only when comparing neutral words with positive words rather than negative words, as here. Thus again there is no immediate reason to doubt the generalisability of the present findings.

References

Arsenault, A., & Dolan, S. (1983). The role of personality, occupation and organization in understanding the relationship between job stress, performance and absenteeism. Journal of Occupational Psychology, 56, 227-240.

Clark, D. M., Salkovskis, P. M., Gelder, M., Koehler, C., Martin, M., Anastasiades, P., Hackmann, A., Middleton, H., & Jeavons, A. (1988). Tests of a cognitive theory of panic. In I. Hand & H.-U. Wittchen (Eds.), Panics and phobias (pp. 149-158). Berlin: Springer-Verlag.

Cooper, C. L., & Marshall, J. (1976). Occupational sources of stress: A review of the literature relating to coronary heart disease and mental health. Journal of Occupational Psychology, 49, 11-28.

Cooper, C. L, Sloan, S. J., & Williams, S. (1988). Occupational stress indicator. Windsor, UK: NFER-Nelson.

Cooper, C. L., Watts, J., Baglioni, A. J., Jr., & Kelly, M. (1988). Occupational stress amongst general practice dentists. Journal of Occupational Psychology, 61, 163-174.

Cox, T., & Mackay, C. J. (1985). The measurement of self-reported stress and arousal. British Journal of Psychology, 76, 183-186.

Dalgleish, T. (1995). Performance on the emotional Stroop task in groups of anxious, expert, and control subjects: A comparison of computer and card presentation formats. Cognition and Emotion, 9, 341-362.

De Houwer, J., & d'Ydewalle, G. (1994). Stroop-like interference in sorting for intrinsic color: A test of the Glaser and Glaser (1989) model. Acta Psychologica, 85, 123-137.

DiMatteo, M. R., Shugars, D. A., & Hays, R. D. (1993). Occupational stress, life stress and mental health among dentists. Journal of Occupational and Organizational Psychology, 66, 153-162.

Fisher, S. (1986). Stress and strategy. Hillsdale, NJ: Erlbaum.

Glaser, W. R., & Glaser, M. O. (1989). Context effects in Stroop-like word and picture processing. Journal of Experimental Psychology: General, 118, 13-42.

Ivancevich, J. M., Matteson, M. T., Freedman, S. M., & Phillips, J. S. (1990). Worksite stress management interventions. American Psychologist, 45, 252-261.

Kucera, H., & Francis, W. N. (1967). Computational analysis of present-day American English. Providence, RI: Brown University Press.

Mackay, C. J., Cox, T., Burrows, G. C., & Lazzerini, A. J. (1978). An inventory for the measurement of self-reported stress and arousal. British Journal of Social and Clinical Psychology, 17, 283-284.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. Psychological Bulletin, 109, 163-203.

Martin, M., Horder, P., & Jones, G. V. (1992). Integral bias in naming of phobia-related words. Cognition and Emotion, 6, 479-486.

Martin, M., & Jones, G. V. (1995). Integral bias in the cognitive processing of emotionally linked pictures. British Journal of Psychology, 86, 419-435.

Martin, M., Williams, R. M., & Clark, D. M. (1991). Does anxiety lead to selective processing of threat-related information? Behaviour Research and Therapy, 29, 147-160.

Mogg, K., Bradley, B. P., & Hallowell, N. (1994). Attentional bias to threat: Roles of trait anxiety, stressful events, and awareness. Quarterly Journal of Experimental Psychology, 47A, 841-864.

Murphy, L. R. (1984). Occupational stress management: A review and appraisal. Journal of Occupational Psychology, 57, 1-15.

Pearlstone, A., Russell, R. J. H., & Wells, P. A. (1994). A re-examination of the stress/illness relationship: How useful is the concept of stress? Personality and Individual Differences, 17, 577-580.

Reynolds, S., Taylor, E., & Shapiro, D. A. (1993). Session impact in stress management training. Journal of Occupational and Organizational Psychology, 66, 99-113.

Tukey, J. W. (1977). Exploratory data analysis. Reading, MA: Addison-Wesley.

Watts, F. N., McKenna, F. P., Sharrock, R., & Trezise, L. (1986). Colour naming of phobia-related words. British Journal of Psychology, 77, 97-108.

Williams, J. M. G., & Broadbent, K. (1986). Distraction by emotional stimuli: Use of a Stroop task with suicide attempters. British Journal of Clinical Psychology, 25, 101-110.

Williams, J. M. G., Watts, F. N., MacLeod, C., & Mathews. A. (1988). Cognitive psychology and emotional disorders. Chichester: Wiley.

A Profile of a Heavy/Problematic Collegiate Drinker: A Literature Review

Michael V. Waggle
St. Mary's University

Abstract

Research has shown that college students drink more alcohol and have more problematic behaviors due to alcohol than the average population. Knowledge of the extremely detrimental effects of alcohol indicates there is a need for some type of intervention early in a student's college career. A literature review would be useful in identifying factors affecting heavy/problematic drinking populations. These factors could be used to develop a profile of a student likely to become a heavy/problematic drinker. Since the decision to drink or abstain is a personal choice, it should be noted that in no way should these factors be used in a cause-and-effect manner, nor should they be seen as an inclusive set. This literature review investigates the correlation of such factors as gender, personality characteristics, drinking expectations, and family background with heavy/problematic drinking. Research shows that males are more likely than females to be heavy/problematic drinkers. Members of Greek organizations, students with family histories of drinking, and those students with histories of deviant behavior prior to the age of 15 were also found to have a high prevalence of heavy/problematic drinking. These profiles could be used to develop a screening program to identify those students with an increased probability of becoming heavy/problematic drinkers.

The statistics on college drinking are truly alarming. Cherry's 1987 study found that approximately 83% of all college students drink, compared to the nationwide average of 63% of the total population. Cohort studies have shown that more students drink as the population moves from freshman through senior years of college (Grodstein, Issaac, Sellers, & Wechsler, 1994). This study showed that frequent-light drinkers have dropped from 14% in 1977 to 1% in 1989, and that 33.3% of men and 47.2% of women that did not drink in their first year of college started to drink in their second year.

College drinking accounts for an estimated $4.2 billion nationwide (cited in Grodstein, et al., 1994). Nystrom, Perasalo, and Salaspuro (1993) found that a mere 10% of the students accounted for more than 42% of the alcohol consumed by collegiate drinkers. Since the college drinking rate is so high, it is important to differentiate those students that drink occasionally and in small amounts from those students that are involved in heavy or problematic drinking. Throughout the research a heavy-drinking student was defined as one who consumed five or more drinks more than once a week. Examples of problematic behaviors included sickness, missing class, damaging relationships, violence, blackouts, or arrest as a result of alcohol consumption. Johnson's 1989 study showed that as drinking levels increased so did problematic behaviors.

The statistics involving heavy/problematic drinkers are just as alarming as the overall collegiate drinking situation. Although the total population of collegiate drinkers has increased, the percentage of heavy drinkers has been fairly stable. Engs and Hanson (1992) cited the proportion of collegiate heavy drinkers as follows: 1982 = 24.4%, 1985 = 24.6%, 1988 = 25.7%, and in 1991 = 26.8%. Although these numbers did not show a statistical significance, it should be noted that they did show a steadily increasing percentage of a growing population. In this population, 90% of all drinkers consumed alcohol in difficult to control environments where heavy drinking was normalized (Andrews, Dana, Kochis, & Pratt, 1993). Barry, Fleming, and MacDonald's 1991 study found 20% of a 1000-student sample were classified as heavy drinkers

while 29% met DSM-III criteria for alcohol abuse. This DSM-III definition for alcohol abuse is inclusive of our operational definition for problematic drinking. Of the 29% of students that met DSM-III criteria, only 1% considered themselves problem drinkers.

Gender

The differences between male and female drinking patterns is one of the most reported variables in alcohol studies. It seems that the gender norms greatly affect the individual's decision to drink and the consequences thereafter. Whereas it has sometimes been accepted for men to drink large amounts of alcohol, it has not been a practice as accepted for females.

Males

In every study researched, males were more likely to be heavy drinkers than were females. Although the drinking rate showed that a comparable percentage of overall men and women drink, 81.1% and 81.6% respectively, men were more than twice as likely to be engaged in heavy drinking (O'Hare, 1990). Nystrom, Perasalo, and Salaspuro (1993) showed that 11.6% of the male students were heavy drinkers, as compared to only 4.9% of the female students. In a two week sampling period, male students drank an average of 30.2 drinks on an average of 5.4 days; whereas female students drank only 16.1 drinks on an average of 4.2 days (Perkins, 1992). Several factors were correlated to men's exorbitant drinking rate. Dennis, Nagoshi, and Wood (1992) found that men involved with heavy drinking showed increased impulsivity and venturesomeness, had an increased perceived norm of drinking, and had more reasons for drinking. Nystrom, Perasalo, & Salaspuro's 1993 study found that 8.9% of the heavy-drinking males exhibited problematic behaviors as compared to 3.7% of the heavy-drinking females.

The problematic behaviors engaged in by men were more likely to involve consequences to other people. When drunk, men were ten times more likely to get into fights than were females and were three-to-four times more likely to drive under the influence (Nystrom, Perasalo, & Salaspuro, 1993). Of the males indicating heavy-drinking patterns, 63% indicated that they have fought while drunk, almost twice that of females (Barry, Fleming, & MacDonald, 1991). In addition to fighting more, heavy-drinking males were more likely to have problems in relationships and with the law, to drink alone, to drink before class, and to have problems at work (Engs & Hanson, 1990).

Females

Although men have been shown to drink more and exhibit more problematic behaviors, the females in these populations may be underrepresented. In all the studies researched, the operational definition for heavy/problematic drinking involved the number of drinks consumed per occasion. These statistics may be underrepresenting females, since females need less alcohol to become drunk. Ksir and Ray (1993) showed that women have a greater percentage of body fat, therefore, less body volume to distribute the alcohol. Women, therefore, generally become more intoxicated on less alcohol. An operational definition that is dependent on blood alcohol level (BAL), not number of drinks, may provide a better representation of the female population.

There was some evidence that the gap between male and female drinkers was narrowing. As women break away from their more traditional roles, it is understandable that they will also develop some of the male counter-productive behaviors. Although males, in general, engage in more problematic behaviors, females report engaging in the same types of behaviors. Perhaps males engage in more of these behaviors simply because more men drink excessively. Perkins (1992) showed that men were only 1.5 times as likely to get hangovers, to miss class, to get behind academically, to have memory loss, or to do something regrettable as were women. This

study also showed that men were only 1.2 times as likely to engage in an action or behavior in which they would not have engaged if sober.

Females were actually more likely to engage in some self-destructive problematic behaviors than were men. For example, the correlation of cocaine use when using alcohol was .49 for women and only .32 for men (Perkins, 1992). This behavior is especially destructive since the combination of cocaine and alcohol makes cocaethylene. This drug combination was second only to the combination of cocaine and alcohol and heroine in causing the most drug related deaths according to the 1991 DAWN data (Ksir and Ray, 1993). Females also reported more suicidal ideations with alcohol use (O'Hare, 1990). Heavy-drinking females may be the result of women trying to break from traditional roles. Johnson (1989) found that, while males were afraid of failure, the heavy-drinking females were more afraid of success. Ksir and Ray (1993) reported that the heavy-drinking females were drinking to relieve tension. Johnson (1989) also found that while most women following traditional roles were concerned with stability and security, most heavy-drinking females were concerned with sensation seeking, were less concerned with success, and were experiencing sex role conflicts.

Other Demographics

Once gender control was initiated, other demographics showed statistical significance with heavy/problematic drinking. The white population showed the most statistically significant results. O'Hare's study (1990) showed a statistical significance for heavy drinking for those students reporting USA/white or English/Scotch/Irish decent ($X^2 = 84.65$). With the heaviest levels of drinking, it is not surprising that the Caucasian population also showed the most problematic behaviors (Curtis et al, 1990). This study showed a mean number of problematic behaviors for Black, Hispanic, and Caucasian men as $X = 2.57$, $X = 3.13$, $X = 6.43$, respectively. This trend was also found for Black, Hispanic, and Caucasian women with $X = 1.86$, $X = 1.92$, and $X = 3.61$, respectively. O'Hare (1990) also found that Catholic and Jewish denominations drank the most while Protestants, other Christians, and "others" showed the least amount of drinking.

Where a student lives has shown statistical significance toward drinking styles. Kayson and Lichtenfeld's 1994 study found that members of a Greek organization had almost twice as many problem behaviors related to drinking as non-members; $X = 7.2$, $X = 4.7$, respectively. This factor showed no significance for students over 35 years of age. It was found, however, that students involved in two or more college organizations (not fraternities or sororities), generally did not show any problematic-drinking behaviors (Cherry, 1987).

O'Hare (1990) showed that as men moved away from campus their likelihood of heavy drinking decreased. Men living on campus were most likely to be heavy drinkers (36.8%). This was almost double the number of heavy-drinking men living independently off campus, 18.2%. Men living at home were most likely to be abstainers. Women, however, showed a different pattern. 21.3% of the women living independently off campus were the heaviest drinkers, while only 13.0% of the women living on campus were heavy drinkers. Finally women living at their parents' home were most likely to abstain, with only 6.0% showing heavy-drinking patterns. The tendency of independent women showing the heaviest drinking patterns supports the theory that women breaking from traditional stereotypes may exhibit heavy/problematic drinking.

Personality Characteristics

There were some personality characteristics that showed statistical significance with alcohol consumption. Gorman & Werch (1988) found that students using external self-control strategies were likely to drink more and, therefore, have more problematic behaviors. This study found that "The self-control strategies that correlated most highly with alcohol-related problems were those aimed at setting time and food constraints, followed by self-reinforcement and

punishment" (Gorman & Werch, 1988, p.32). Deviant behavior was another personality characteristic correlated with alcohol abuse. A study by Barry, Fleming, and MacDonald (1991) found that "students with a history of deviant behavior prior to age 15 had a 1.87 relative probability of being diagnosed for alcohol abuse, an almost twofold risk" (p.444).

Expectations

The drinker's expectations of alcohol has been shown to be a better predictive factor than demographic or background data (Dennis, Nagoshi, & Wood, 1992). This study showed that "the expectancies of social and physical pleasure and of tension reduction contributed the greatest amount of predictive power with regard to. . . frequent, problematic drinking styles" (p. 473). Heavy drinkers were most likely to drink alcohol for the effect of disinhibition. Heavy drinkers were also more likely to think that hostility, depression, and impairment would not be seen from other people as a negative side effect (Dennis, Nagoshi, & Wood, 1992). Dunlosky and Leavy (1989) indicated that as consumption increased, the drinkers' definition of problematic drinking became more liberal.

Reasons for Drinking

Dennis, Nagoshi, & Wood (1992) found that light and moderate drinkers generally drank for "celebratory" reasons. These reasons included parties, or when feeling sociable. Heavy drinkers, on the other hand, drank for "pathological" reasons. These pathological reasons were generally self-medicating situations when tense, when angry or irritable, or when the drinker wished they were another person. The heavy drinker seems to have found that drinking can relieve negative mood states.

First Intoxication

Grodstein, et al., (1994) showed that once drinking behaviors started, they generally continued. 97% of those students polled who drank during their freshman year continued to drink throughout their sophomore year. Another factor significantly correlated with heavy drinking was the age at which the person initially started drinking.

Wechsler and McFadden (1979) stated that the best predictors of heavy drinking in college was heavy drinking in high school (cited in Grodstein, et al., 1994). Andrews, et al., 1993 study showed that the earlier a person begins drinking, the greater the number of problems, number of drinks consumed, and number of reasons for drinking. Grodstein, et al., 1994 study also showed that if students did not begin to drink until college, they were much more likely to be light to moderate drinkers throughout college. If the drinking behavior was "learned" in high school, however, it was more likely to be heavy and problematic.

Family Background

A complete discussion of the correlation of family background and alcohol abuse goes well beyond the scope of this paper. Entire volumes have been written on this topic. This paper will focus on some of the more global theories involving alcoholic families.

A review of some studies showed somewhat mixed findings. Kayson and Lichtenfeld (1994) found that students with alcohol-related problems in their nuclear families showed a statistically higher mean of problems. Other studies have not been so conclusive. A 1993 study by Andrews et al. could not find significant results for family backgrounds compared to number of drinks consumed or number of problematic behaviors. Although the literature is somewhat mixed, disregarding family background would be counterproductive. It may be that the statistical

measures could not filter out the family background factors since such a large population of college students drink.

Conclusion

Although men are more likely to be involved in heavy/problematic drinking, there are many factors that are identical in the profile for male or female heavy/problematic drinkers. The heavy/problematic drinker is most likely to be involved with a Greek organization, be under the age of 35, Caucasian, and Catholic or Jewish. They are more likely to have some history of alcoholic problems in their family backgrounds. They generally started drinking heavily in high school and have a history of deviant behaviors prior to the age of 15. They are also more likely to have lenient attitudes towards drinking and drink for self-medicating reasons.

There are some differences between male and female heavy/problematic student drinkers. Generally males follow more traditional male-type behaviors; whereas, females follow more untraditional roles. Female heavy/problematic student drinkers are more likely to live independently off campus and be concerned about success, not failure. Male heavy/problematic student drinkers, however, are more likely to live on campus and be concerned about failure.

References

Andrews, W. M., Dana, R. Q., Kochis, R. A., & Pratt, P. A. (1993). Problematic college drinking behaviors as a function of first intoxication. Journal of Alcohol & Drug Education, 38(2), 92-99.

Barry, K. L., Fleming, M. F., & MacDonald, R. (1991). Risk factors associated with alcohol abuse in college students. American Journal of Drug & Alcohol Abuse, 17(4), 439-449.

Cherry, A. (1987). Undergraduate alcohol misuse: Suggested strategies for prevention and early detection. Journal of Alcohol & Drug Education, 32(3), 1-6.

Dennis, D. A., Nagoshi, C. T., & Wood, M. D. (1992). Alcohol norms and expectations as predictors of alcohol use and problems in a college student sample. American Journal of Drug & Alcohol Abuse, 18(4), 461-476.

Dunlosky, J. T., & Leavy, R. L. (1989). Undergraduate student and faculty perceptions of problem drinking. Journal of Studies on Alcohol, 50(2), 101-107.

Engs, R. C., & Hanson, D. J. (1990). Gender differences in drinking patterns and problems among college students: A review of the literature. Journal of Alcohol & Drug Education, 35(2), 36-47.

Engs, R. C., & Hanson, D. J. (1992). College students' drinking problems: A national study, 1982-1991. Psychological Reports, 71(1), 39-42.

Gorman, D. R., & Werch, C. E. (1988). Relationship between self-control and alcohol consumption patterns and problems of college students. Journal of Studies on Alcohol, 49(1), 30-37.

Grodstein, F., Isaac, N. E., Sellers, D. E., & Wechsler, H. (1994). Continuation and initiation of alcohol use from the first to the second year of college. Journal of Studies on Alcohol, 55(1), 41-45.

Johnson, P. B. (1989). Personality correlates of heavy and light drinking female college students. Journal of Alcohol & Drug Education, 34(2), 33-37.

Kayson, W. A., & Lichtenfeld, M. (1994). Factors in college students' drinking. Psychological Reports, 74(3, Pt 1), 927-930.

Ksir, C., & Oakley, R. (1993). Drugs, Society, & Human Behavior (6th ed.). St. Louis: Mosby. (pp 179-226).

Nystrom, M., Perasalo, J., & Salaspuro, M. (1993). Screening for heavy drinking and alcohol-related problems in young university students: The CAGE, the Mm-MAST and the trauma score questionnaires. Journal of Studies on Alcohol, 54(5), 528-533.

O'Hare, T. M. (1990). Drinking in college: Consumption patterns, problems, sex differences and legal drinking age. Journal of Studies on Alcohol, 51(6), 536-541.

Perkins, H. W. (1992). Gender patterns in consequences of collegiate alcohol abuse: A 10-yr study of trends in an undergraduate population. Journal of Studies on Alcohol, 53(5), 458-462.

Military Psychologist: What is Military Psychology?[1]

Martin F. Wiskoff
BDM International, Inc.
Major Dana H. Lindsley
Office of the Assistant Secretary of Defense
(Force Management Policy)

Introduction

> The Handbook of Military Psychology (Gal & Mangelsdorff, 1991) defines military psychology as the "application of psychological principles, theories and methods, within a military environment." Driskell and Olmstead (1989) in their review of psychology and the military state that "the field of military psychology is defined neither by a common set of techniques (as is experimental psychology) nor by a common set of problems (as is developmental psychology) but rather by the area or context of application—the military."

Many may not be aware that there is a discipline of military psychology. In fact, the Division of Military Psychology was one of the original 19 divisions created within the American Psychological Association in 1946.

Military psychology is a microcosm of psychology and consequently offers opportunities to psychologists of all persuasions, including those who wish to spend their career or a portion of it in a military uniform. It is also a discipline that crosses international boundaries, with military psychologists found in many countries. The problems addressed by the research are of concern to the militaries of all nations and there is the potential for cross-national research efforts, technical exchanges, and even assignments to serve jointly with military forces of other nations.

Perhaps of greatest importance, military psychology offers the opportunity to make a significant difference in the lives of individuals and in the stability of our nation. A small sample of the types of contributions that can be made by military psychologists include (a) working in mental health or family counseling clinics to improve the lives of service personnel and their families; (b) performing research on the effects of battlefield environmental factors on soldiers in order to prevent or reduce battlefield casualties; and (c) analyses of humanitarian and peacekeeping missions to determine procedures that could save military and civilian lives.

A brief review of the proud history of U.S. military psychology provides a foundation for examining the range of jobs and settings in which civilian and uniformed military psychologists are found.

History of Military Psychology

The psychology of war has been studied by military tacticians for as long as human beings have waged battles. Success on the battlefront is dependent on behavioral issues, such as assessment of troop readiness and knowledge of an opponent's vulnerabilities, as often as on the actual size of the opposing forces.

In the years leading up to World War I, psychology had begun to emerge as a field of scientific study and application. American psychologists had become intrigued with the mental measurement work of Dr. Binet in France and with the scientific management movement to enhance worker productivity. However, it was the problems of assimilating millions of U. S. civilians into the armed services that brought the tools of psychologists to the military environment and created the discipline of military psychology in the U.S.

At the start of the war, a group of psychologists headed by the president of the American Psychological Association, Dr. Robert M. Yerkes, met to discuss how psychology could assist in the war effort. The successful program of mental testing of recruits with the Army Alpha and Beta examinations, which resulted in the appropriate placement of new soldiers into military jobs and officer training, is indelibly identified as the genesis of military psychology. It also served as the subsequent model for group intelligence testing for both military and civilian application.

In addition, during the short time frame between U.S. entry in 1917 until shortly after the war in 1918, psychologists addressed many other issues: measurement of troop morale and assimilation into the military; development of special trade tests to assess skills, such as combat leadership or flying aptitude; assessment of emotional instability; and measurement of human performance. Immediately after the war, psychologists conducted surveys to assess the attitudes of soldiers, including their opinions about their own military service. Psychologists who contributed during World War I—truly the first military psychologists—included such notables as Edwin G. Boring, James McKeen Cattell, G. Stanley Hall, Walter Dill Scott, Carl E. Seashore, Edward K. Strong, Lewis M. Terman, Edward L. Thorndike, John B. Watson, and Robert S. Woodworth.

There was a hiatus in the study and practice of military psychology during the 1920s and 1930s, but at the start of World War II the military reestablished a psychological research program during which more than 2,000 psychologists, civilian and uniformed, would address military problems. Military psychology, born in the first world war, matured in World War II. Former areas of inquiry were revisited, and many new ones were added: military leadership; the effects of environmental factors on human performance; military intelligence; psychological operations and warfare; selection for special duties; and the influences of personal background, attitudes, and the work group on soldier motivation and morale.

Military psychology was the dominant theme in psychology during World War II. As reported by Driskell and Olmstead (1989), "In 1943, fully half of the pages of the Psychological Bulletin were devoted to topics of military psychology, and from 1943 to 1945, one in every four

psychologists in the country was engaged in military psychology." After the war, much of what had been learned found ready application in other public and private sector settings.

In the 50 years since the World War II military psychological research has continued its tradition of innovation and has provided leadership to the civilian sector. Military service research laboratories were created, and extensive programs were established to fund research at universities and by private contractors. In addition, military psychologists have participated in large social policy programs conducted in the military that were designed to increase diversity and equal opportunity. These programs include integrating racial and ethnic groups, eliminating sexual discrimination and harassment, employing women in combat and in work settings designed for men, utilizating low capability recruits and rehabilitating juvenile delinquents, drug testing, psychological treatment for personal lifestyle problems, and smoking abatement in the workplace. Military psychologists had the opportunity to research, evaluate, and make national policy recommendations concerning these programs.

The Department of Defense (DoD) employs more psychologists than any other organization or company in the world. The downsizing of the military in the 1990s has been accompanied by a corresponding reduction in research and psychological support to the operating forces. The future of military psychology is assured, however, as long as there is a need for troops to defend our country and perform peacekeeping missions around the world.

Types of Work Pursued in Military Psychology

The Handbook of Military Psychology (Gal & Mangelsdorff, 1991) is the most comprehensive single source of information concerning the types of work performed by military psychologists. To assist the reader in relating this discussion to the Handbook, Table 1 displays its seven major categories of military psychology (slightly modified) and two additional areas (indicated by an asterisk) that the Handbook covers only minimally. Because these nine application areas all have the same goal of improving the performance and adjustment of personnel within the military environment, the actual work conducted across the areas overlaps somewhat.

There is another "type-of-work" dimension that may prove useful to keep in mind while reviewing these nine areas. Military research is funded within discrete categories on a dimension that ranges from basic to applied. The goal of basic research is to develop new knowledge or technologies with potential application to military problems. The more applied types of research seek to explore and evaluate the utility of new technologies in operational military environments. This often includes developing prototype systems (e.g., computerized performance measurement) and conducting feasibility testing with service personnel. An additional class of applied work involves conducting studies (e.g., surveys, database analyses) that provide management and policymakers with information on which to base policy decisions (e.g., whether to revise enlistment incentive programs). Table 1 lists the nine areas of military psychology along with the most closely related psychological disciplines.

Table 1 Types of Work Within Military Psychology

Military Application Area	Most Closely Related Psychological Disciplines
Selection, Classification, and Assignment	Evaluation and Measurement, Cognitive, Industrial/Organizational
Training and Education	Experimental, Cognitive, Educational
Human Factors Engineering	Human Factors, Ergonomics, Experimental
Environmental Stressors and Military Performance	Physiological, Psychopharmacological, Psychobiology, Experimental
Military Leadership and Team Effectiveness	Social, Industrial/Organizational
Individual and Group Behavior	Personality, Social, Adult Development
Clinical and Consulting	Clinical, Counseling, Consulting, Family and Health, Community
Manpower Management and Decision Making Support	Advertising; Evaluation and Measurement, Social, Industrial/Organizational
Special Subjects and Situations	Psychology of Women, Study of Social Issues, Peace, Personality, Health, Clinical

Work Settings For Military Psychologists

Military psychologists often work in a broader range of settings than would be the case for most other psychological disciplines. Because of the large number of bases, schools, offices and other sites under military jurisdiction, however, there are opportunities for assignment at many different locations in the U.S. and abroad. Temporary assignments to serve the troops in combat zones, develop studies, collect data, present research findings, etc., are commonplace. Table 2 displays the six major types of settings in which military psychologists are located.

Table 2 Military Psychology Work Settings

Major Settings	Examples of Locations
Research facilities	Military laboratories and field units; contractor offices
Educational facilities	Colleges and universities; military educational institutions
Medical centers, hospitals and clinics	Military hospitals; outpatient clinics; mental health centers; drug treatment centers; prisons
Military schools and bases	Service training schools; military bases in the U. S.
Military deployments overseas	Military overseas bases and small missions; combat zones; military hospitals
Military organization offices	The Pentagon; service headquarters commands

Professional Linkages

Military psychologists have the opportunity to join national and local professional organizations that reflect their specific research interests and publish the results of their research in a host of journals that cover the diverse areas of interest within the field. The primary identification for many military clinicians and researchers is the Division 19—the Division of Military Psychology—of the American Psychological Association (APA). It is common for military psychologists to belong to other APA divisions, such as Experimental Psychology (Division 3); Evaluation, Measurement and Statistics (Division 5): Clinical (Division 12); the Society for Industrial and Organizational Psychologists (Division 14); Applied Experimental and Engineering (Division 21); Health (Division 38); and Family (Division 43). Many are members of other professional organizations such as the American Psychological Society, Human Factors Society, and the Inter-University Seminar.

The Division of Military Psychology publishes its own quarterly journal, Military Psychology, which features original behavioral science research and scholarly integration of research findings performed in a military setting. Military Psychology has published contributions from a number of countries and has featured special issues on topics of particular interest to the military research community: Team Processes, Training and Performance; Women in the Navy; Stimulants to Ameliorate Sleep Loss During Sustained Operations; and Military Service and the Life-Course Perspective. Other special issues scheduled for publication in 1996 and 1997 include Military Occupational Analysis; The Impact of Chemical Protective Clothing on Performance; and Enhanced Computerized Adaptive Testing.

Applications

Given the military's need and penchant for the most current battlefield and management technologies, much military research is at the cutting edge of science. Military laboratories offer a psychologist the unique opportunity to conduct research without collateral requirements to teach or consult. Laboratory personnel can establish a career path to include increasing research management responsibilities and possible service in decision making roles within government. Uniformed psychologists have unusual opportunities to perform research, to provide clinical services in a unique environment, or to consult on matters of international importance. Most military research has important applications in the private sector as well. Joint government-industry undertakings are becoming commonplace. Military issues and technologies cross national boundaries, and the international community of military researchers shares information at military and professional conferences and during exchange visits which helps keep the profession vital in the U.S. and abroad.

References

Driskell, J. E., & Olmstead, B. (1989). Psychology and the military: Research applications and trends. American Psychologist, 44, 43-54.

Gal, R., & Mangelsdorff, A. D. (Eds). (1991). Handbook of military psychology. New York: Wiley.

Are We Winning the War on Drugs?[1]

Colonel Clark Hosmer, Ph.D.
Shalimar, FL 32579

Abstract

The increased use of drugs and costs in money and civil rights suggest we are not winning the war on drugs. Some positive domestic programs and some overseas activities, however, are promising. But the war policy is steady on. Politicians cater to the public's belief that drugs are evil. The Base Realignment and Closure (BRAC) procedure eases a politicians' problem of loss of military bases. Use of the procedure could be a solution to our problems with drugs.

In 1980, President Ronald Reagan and Dr. William Bennett as the Drug Czar, launched a crusade called the War on Drugs. As was true of the European Crusades, enormous resources of people and materiel have been invested in the war on drugs. Are We Winning the War?

The most desired measure of winning would be less use of drugs. The National Household Survey on Drug Abuse, however, reported that use of drugs in the United States is the highest rate in the industrial world. Further, the University of Michigan's "Monitoring the Future" reported the use of marijuana by high school youngsters nearly doubled from 1992 to 1995. Police estimate that as much as nine pounds of drugs come through our borders for every pound intercepted. On January 3, 1996, James Milford, special agent in charge of the South Florida branch of the Drug Enforcement Agency (DEA), briefed the Greater Miami Chamber of Commerce. He announced that Miami has become the Columbian drug empire's North American headquarters. He reported the Cali drug cartel is becoming increasingly difficult to stop. The cartel, he said, earned some $8 billion last year on which we collected no taxes. A California police officer said the cartel does not bother to count money. They sort and weigh it.

From spending a billion dollars per year at first, the direct cost of this war has risen to more than $14 billion per year. One side effect is the cost of confinement of prisoners the majority of whom now are convicted of drug-related crimes. We have the highest national rate of convicts in the Western world. At an estimated $30,000 per year per prisoner, our federal, state, and local confinement costs total $45 billion per year and are increasing. We have been forced by the war to build more prisons and jails.

A different, additional cost has been in civil rights. The Supreme Court approved in 1982 the warrantless search of a car, a brief case in it, and the glove compartment; in 1983 a boat; in 1984 fenced private property; in 1985 a purse; and in 1987 a house and personal papers. The discovery of any amount of an illegal drug is used to justify seizure of money, boats, cars, homes, and farms, whether or not the owner personally uses or pushes drugs or is charged with a crime. On January 12, 1996, the Supreme Court agreed to decide whether the easy confiscation of property under civil law may be used under criminal law. Lower courts have held that under criminal law easy

confiscation is unconstitutional double jeopardy. Is it not strange that police may use easy confiscation under civil(!) law?

In sum, increased drug use and costs in money and in civil rights suggest we are losing the war on drugs. Moreover, the Merriam-Webster's Collegiate Dictionary (1994) says that arbitrary exercise of repressive control by police and the legal deprivation of basic civil rights are characteristics of a police-state.

Is There No Good News?

Three television programs reported problems with the war on drugs. But they also reported good news on activities that show promise, both here and abroad. .

In August 1994, MTV broadcast "Straight Dope." On April 6, 1995 ABC's Ted Koppel and Catherine Crier presented "America's War on Drugs: Searching for Solutions" and again on Nightline, "What's the Best Way to Win the War on Drugs?" On June 20, 1995, CBS's Walter Cronkite presented "The Drug Dilemma: War or Peace?"

Cronkite alleged that the war on drugs is the most destructive social experiment this nation has ever tried. He cited Steven Duke, Yale professor of Criminal Law, in his "America's Longest War: Rethinking Our Tragic War against Drugs." Duke says the war on drugs does not lessen but creates crime, destroys education in inner cities, packs prisons, and violates our 4th, 5th, and 6th Amendments.

Cronkite, Koppel, and Crier featured "harm reduction" as a better strategy than prohibition. The idea is to face the fact that around the world humans like whatever they associate with relaxation and feeling good. Many humans class coffee, alcohol, tobacco, and drugs as helping them to relax and feel good. The United States tried prohibition of alcohol. That was rejected. Now we attack abuse of alcohol as a problem that, bad as it is, is less harmful to our culture than Prohibition was. Prohibition brought the evils of abuse of bootleg liquor, disparagement of national law, and untaxed wealth for gangsters. Put drugs for liquor and that describes today.

Cronkite reported on a scatter of promising programs. Brooklyn Judge Rose McBrien created "Drug Treatment Alternative to Prison" (DTAP). The DTAP two year treatment produces re-arrest rates less than half of non-treated prisoners, 16% vs. 40%. It costs $18,000 vs. $44,000 per prisoner per year. Cultural benefits and savings, too. Of course, selection of those treated may be a factor.

The Amity program in Arizona treats drug-user mothers and their children for $17,000 per convict per year. Payoff in cures looks promising, but more time for long-term data is needed.

Indiana tried "Indiana Students Taught Awareness and Resistance" (I-STAR). The children who had I-STAR have lower rates of drug, alcohol and tobacco use. Congress, however, has cut money for school based programs like I-STAR.

The Federal Centers for Disease Control reported that without increasing use of drugs, every time clean needle exchanges have been tried, they have dramatically reduced the spread of HIV.

On whether decriminalization of illegal drugs would produce an explosion in their use, Cronkite reported that after an initial drop in rate of use of alcohol under the 18th Amendment, use of bootleg alcohol gradually rose to two-thirds of pre-prohibition level. After repeal, the rate of use of alcohol rose quickly, but only to about the pre-prohibition level. The pre-prohibition level of the use of drugs was not headline news and has not been reported.

Cronkite reported that Targeting Systems found that if cocaine were legal, only 0.9% said they would try it and only 4.2% said they would try marijuana. Those rates are less than those who now use cocaine and marijuana.

RAND corporation reported that investment in treatment reduces the use of drugs as much as seven times as large an investment in enforcement and punishment.

Overseas

Unlike the United States, Holland invests heavily in treatment. They do not bother drug users and have proposed legalizing domestically grown marijuana to compete with imported varieties. Less than 2% of the resident population use marijuana and less than 0.2% use hard drugs. Those are the lowest rates in the European nations.

Spain and Italy decriminalized marijuana. Then Italy reversed that policy. Germany and Switzerland are trying modifications of their antidrug policies, but strong forces oppose change. Australia has an official proposal for political review of trying availability of heroin in the capital district.

Dr. John Marks, near Liverpool, England, provided free drugs and clean needles to resident drug users who attended weekly meetings. During the first eight years of his clinic, the rate of drug users dropped from 0.2% to 0.01%. The drop in thefts and burglaries impressed the police, insurance companies, and merchants, one of whom sponsored an international conference to discuss the program.

Marks said both prohibition and its opposite --- legal availability plus pushing by commercial ads as we have for alcohol --- are wrongheaded. Both extremes of 1) prohibition and 2) alluring commercials lead to excessive use of problem substances. His suggestions are control the availability and ban advertisement of problem substances.

As an example of the state of flux about drugs in Europe, after eight years of the compelling benefits produced bythe Marks clinic, opponents of his controlled availability of drugs were encouraged by United States authorities to close Marks' program. It was closed.

This overview suggests we are not winning the war on drugs. We are creating more problems than we are solving. Other nations vary in their groping with the drugs problem. Our government is holding to the course of the war on drugs.

Why do Politicians Push the War on Drugs?

President Reagan and Drug Czar Bennett did a thorough job in convincing the public that the problem is moral: drugs and drug users are evil. Politicians find that when they show a hard attitude toward drugs, their constituents applaud. To keep their jobs, politicians need to comply with the line of righteousness.

Please note that people tend to believe leaders who label problem things and actions as "evil." Believers are an iron opposition to politicians who may question the validity of the "evils." If the leaders were content with saying only that the problem things and actions will produce bad consequences for people, the problem things and actions would be more nearly subject to objective analysis and solution. But beliefs in evils are facts. They are one of our human characteristics that are givens and, as such, are dealt with elsewhere.

Cronkite recommended having a bipartisan commission analyze our drug problems and present a comprehensive drug policy for the future. That sounds good.

The procedure used in the Base Realignment and Closure (BRAC) program has succeeded in relieving the individual members of Congress from personal responsibility for loss of military facilities with their Federal jobs and dollars. Under the requirement that the Congress and the President must accept the whole of the recommended program without red-lining particular items, each politician can say he fought but lost at the level of the Commission's considerations.

Conclusion

Perhaps now is time to update the analysis of the Shafer Commission of 1972. It recommended reversal of the prohibition of drugs. A new bipartisan Commission under BRAC-type procedures could illuminate the complex of our problems with all drugs. Their recommendations, if adopted, probably would improve our fiscal, civil, and cultural health.

---

[1] For their criticisms, I thank Calvin Bass, Kevin Casey, T. Bruce Graham, Terry Hecker, James and Ruth Ann Hickey, Gary Hosmer, Lynn Karsok, Bruce Netschert, and Otha Spencer.

References

Duke, S., (1994). America's Longest War: Rethinking our Tragic Crusade Against Drugs. New York: Putnam Publishing Group.

Merriam-Webster's Collegiate Dictionary. Tenth Edition, (1994). Springfield, MA. Merriam-Webster, Inc.

Using Existential and Cognitive-Behavioral Techniques in the Design of a Short-Term Therapy Group for Incest Survivors

Timothy P. Kopania
St. Mary's University

Abstract

Reports of incest in our society have risen sharply over the past few decades. Increasing health care costs prevent many survivors from obtaining long-term individual and group therapy. This paper presents a theoretical framework for therapists to use when developing short-term therapy groupsfor incest survivors. This prescriptive model is a stepwise approach foundedin existential and cognitive-behavioral techniques. Factors considered in the development of the group are the type(s) of psychological symptoms and theco-occurrence of those disorders, the type of contact experienced by the survivor (coital vs noncoital), group homogeneity, and family interaction.

Incest is a long-term problem in our society, the implications and prevalence of which have only come to light recently. An article by Joy (1987) states that where the chances of harming a child's normal developmental growth are concerned, incest ranks higher than abandonment, neglect, physical maltreatment, or any other form of abuse. The most common developmental problem is lowered self-esteem, possibly resulting from internalized guilt and self-blame (Morrow & Sorell, 1989).

There has been much research performed to model the recovery process of incest survivors. It seems most reports of incest do not occur until many years after the initial acts. Therefore, secrecy plays a very large role in the incest survivor's life (Josephson & Fong-Beyette, 1987; Singer, 1987). Joy (1987) developed a four-stage approach to help allow therapists access to deeper issues involved with incest survivors. The first stage is providing the client with the opportunity for emotional catharsis within a safe and accepting environment. Establishing trust and rapport between the client and therapist are vital during this first stage. The second stage involves the promotion of past experience with current life and an understanding of family and personal dynamics which may have facilitated the occurrences of incest. It is imperative at this point to uncover specific information about the incestuous situation. The third stage is to help the client develop positive self-feelings such as self-acceptance and appreciation of personal strengths. The final stage deals with issues in expanding the ability to trust and relate intimately with others.

Josephson and Fong-Beyette (1987) describe a somewhat different model. Their first stage deals more with the client's readiness to admit incestuous acts occurred at all. The research suggests some clients will reveal experiences during counseling and will eventually resolve conflicts associated with their victimization, but others will avoid revealing their histories even with many years of counseling. Many of the latter group will never return for counseling after finally disclosing they actually were victims of incest. This step is essentially the second stage -- the decision by the client to terminate or continue into the deeper issues of denial and outward adjustment. The final step is an integration and resolution stage where the survivor is forced to deal with how the past is affecting them today. This step not only involves the survivor's readiness and ability to recover, but the counselor's ability in direct questioning, specific characteristics (i.e., are the therapist and the perpetrator the same gender?), and positive counselor reactions to initial disclosure.

The psychological symptoms that appear with incest have many diverse characteristics (Morrow & Sorell, 1989). These include low self-esteem, substance abuse, self-destructive

behavior, difficulties in interpersonal relationships, sexually related issues, eating disorders, delusions, chronic depression, post-traumatic stress disorder, and trust issues (Singer, 1989; Sparks & Goldberg, 1994). The co-occurrence of these disorders along with the survivor's tendency towards secrecy may explain why most incest cases are not uncovered for years.

Thorpe and Olson (1990) have hypothesized about how the anxieties and fears associated with incest are maintained. This theory, associated with information processing, discusses how the memories of trauma may be stored in networks that contain stimuli, responses, and interpretations. Stimuli vary from certain places, to time of day, and the similarities of strangers to a past perpetrator. A schema develops to the point where any stimuli associated with the fear network cause an incest response which may be fear, anxiety, and/or feelings of helplessness. Therefore, the victim becomes predisposed to attend to stimuli that fit the schema and ignore stimuli that do not fit.

Another large factor in the type of psychological problems developed after incest is the form(s) of sexual contact experienced by the survivor. A study by Morrow and Sorell (1989) predicted the association of the types of sexual contact with self-esteem, level of depression, and negative behaviors. They found that coital contact was associated with lower self-esteem, higher levels of depression, and greater numbers of antisocial and self-injurious behaviors than other types of noncoital incestuous contact. In addition, they found the "severity of abuse was the single most powerful predictor of self-esteem, depression, and negative behavior" (p. 683).

The statistics indicating the prevalence of incest in our society can be quite surprising. Josephson and Fong-Beyette (1987) cite "it is estimated that 9% to 16% of all women experience incest before age 18" (p. 475). Joy (1987) stated the figure may be as high as 25% of all women. In a study of 18- to 22-year old female college students, only 2% to 5% of those seeking counseling reported a prior incestuous relationship. The difference between the prevalence in society and those seeking counseling suggests either a repression of the experience or a lack of faith in counseling services. Male survivors of incest are largely underrepresented in literature. Singer (1989) reported that between 11% and 47% of the total number of survivors are male, although other researchers have found varying statistics.

Based on the statistics found in current research, there is a tangible need for specific programs to be developed for survivors of incest. Group therapy is a definite avenue for therapists to take because of its efficiency and economic aspects in today's environment of rising health care costs.

Aside from being efficient and economically attractive, group therapy offers other benefits over individual therapy as well. Sparks and Goldberg (1994) showed many incest survivors have trouble with interpersonal relationships and have a feeling of being totally alone in their suffering and recovery. Group therapy presents a tool for the survivors to use in order to see others with similar experiences and feelings towards their perpetrators, interpersonal relationships, etc. (Singer, 1989). In addition, the lack of most families to admit there ever was a problem can leave the survivor feeling outcast and alone, often without any validation from siblings, parents, or relatives. Again, group therapy can provide a platform to discuss these commonalties between survivors.

Sparks and Goldberg (1994) report group members felt empowered when they realized that by coming to a group and discussing their stories, they were changing their social reality. The group experience can be used to overcome fear of the perpetrator, isolation, shame, and general feelings of guilt over disclosing the incestuous experiences in the first place. In general, a positive, corrective experience is offered, with the focus being on specific goals and to experience growth through identifying with other incest survivors (Sparks & Goldberg).

There are several other factors that should be considered in the formation of a group. As with most groups, the vast majority of the members will most likely come from referrals, but should be on a voluntary basis. Because of this, members should be asked to make a commitment to the group after being fully informed of its purpose, members' rights, etc. The number of members in the group should be limited to between six and ten, and the gender should be either strictly male or female because men and women often suffer from different issues (Singer, 1989). The group should also be homogeneous with respect to age because "older and younger women have different developmental tasks and have not experienced the same degree of negative social stigma" (Sparks & Goldberg, 1994, p. 147). The group should meet at a time which will interfere as little as possible with daily life, usually late afternoon or early evening. The group leader should be the same gender as the group so as not to cause any negative reactions from group members because of spurious similarities between a male therapist and perpetrator. Furthermore, the group should not meet in any locations where fear responses may occur due to environmental stimuli. To be able to concentrate as much as possible on specific issues, a therapist should consider limiting membership to survivors who experienced only coital or noncoital contact.

The following paragraphs are designed to provide a prescriptive model for therapists to use with incest survivors. This stepwise approach is based on current research and should be used as a guideline in developing a group using existential and cognitive-behavioral approaches.

This short-term group should meet for between six and ten sessions at approximately two hours each, once a week. The first half of each session should be devoted to a guided discussion of prearranged topics. This existential discussion focuses on Gestalt-like issues of helping members to discover their own feelings about their experiences with incest and to show the other members they are not alone in their victimization. After this, a fifteen minute break is given. The second half of each group will be devoted to cognitive-behavioral techniques aimed at managing fear responses and Gestalt letter writing techniques.

The group should have both long- and short-term goals. The short-term goals for the group will be to develop trust among the group members and a sense of rapport with the therapist. By doing this, the members will feel more comfortable disclosing their past experiences and current feelings about those experiences with the therapist and each other. The long-term therapy goal for this group is to develop the tools to change negative behaviors and eventually self-esteem through existential and cognitive-behavioral techniques. Teaching about cognitive restructuring and systematic desensitization will prepare the survivors to eventually deal with their issues outside of the group experience (Sparks & Goldberg, 1994).

Session 1

The first session should begin by giving the members some evaluative measures to assess their levels of co-occurrence, ability, and qualifications for the group. The prospective members should be given the Beck Depression Inventory to measure their level of depression, the Unwanted Sexual Events Scale to measure the depth and type of childhood sexual abuse, and the Family Environment Scale to measure the survivor's perceptions of their family's interactions, organization, decision-making strategies, and allowance for personal growth.

In this first session, the leader should introduce the purpose of the group, discuss their availability to members, explain the confidentiality policy, and set the ground rules for group membership. During this time the members should also determine the goals of the group. The leader may suggest to have an open invitation to all the members to tell the group whatever they want the other members to know. By beginning in this manner, using time constraints, and focusing on specific goals, a sense of security is provided to the group members. This allows "members to bond and take risks in revealing information about themselves very quickly" (Sparks & Goldberg, 1994, p. 142). Also, the leader may discuss the benefits the group can have on the

individual. By sharing the responsibility, the survivor is involved in making the recovery process meaningful. Another effective opening technique is to talk about the difference in meaning between a 'survivor' and a 'victim'.

Session 2

The second session should begin with a "go-around" of the members' feelings about the first meeting. The guided discussion of this meeting focuses on the schemata of the survivors. They can be asked to talk about how they feel their lives have changed because of their victimization as a child. The focus is on issues of security and negative behaviors they have today. This is a good opportunity to stress that the group must have an atmosphere of openness without any judging or criticizing.

After the break, the group discusses the Gestalt technique of letter writing. These are letters which are written to the perpetrator, never meaning to be sent. A letter writing exercise can be effective for ventilation and clarification (Singer, 1989). The members should be asked to imagine what they would write, and potential reactions of their abuser to anticipate possible responses. Images of the abuser crying or asking forgiveness can help the survivors restore their feelings of control. Also, the possibility of less positive responses should be discussed and what those might be. Towards the end of this session, the leader should ask the members to give some serious thought as to what they might actually write, and to think about the possibility of doing it during the next session. It is important to note here that the purpose of this exercise is not to relive the incestuous experience with full affect which will simply re-traumatize the survivor. Issues such as these are inappropriate for short-term groups to attempt to accomplish.

The last fifteen minutes of this session and each session henceforth should be devoted to developing a sense of closure among the members. The leader stops the discussion and/or exercise and ask if anyone has any feelings they have not had the opportunity to share with the group. During this time, the leader should pay particular attention to the discussion to develop topics for the next session's opening discussion.

Session 3

The third session should begin with a discussion about anxiety. Specifically, the type of stimuli that produce anxiety and evoke other types of negative behavior. By identifying these types of stimuli, members can find what they have in common and learn how each of them deal with those stimuli, if at all. The focus will then be on alternative strategies members can use to change their reactions to the identified stimuli.

The second part of the therapy during this meeting focuses on members actually writing letters to their perpetrators. The leader should stress that the letters not be 'sugarcoated', and they do not need to go into much detail. The letters will tell how the members feel about the abuse to ensure any misgivings the perpetrator may have had about the abuse are dispelled, such as the survivor actually enjoying the abuse, etc. The end of the letter will be focused on what the survivor wants the perpetrator to do (apology, restitution, conditions for future contact) and feelings towards that person (Singer, 1989).

During the closure time, the leader should remind the group that their time together is now half over. The members can begin to think about areas where they still want to go, and their feelings about the group so far. The members should also focus on remaining goals they may wish to pursue.

Session 4

This discussion should begin with a "go-around" of the members' feelings towards the discussion of stimuli from the last meeting. The focus should be on issues of security and negative behaviors. It is important for the group leader at this point to steer away from phrases the members may tend to use like "my stolen childhood" since they are merely euphemisms and do not accurately portray how the members are actually feeling. Time should be spent on exploring how the members make themselves feel secure, and how those methods may be applied to a wider range of situations to enhance the member's quality of life.

During the second portion of this session, the leader should introduce the concept of cognitive restructuring, particularly self-instructional training. This procedure has, essentially, three phases (Thorpe & Olson, 1990). The first is educational, where the leader gives an explanation of the role unhelpful thinking patterns have on producing and maintaining negative behaviors. The second phase is the rehearsal phase. Here the members practice positive, coping self-statements to help with difficult stimuli. The application phase is where the members actually practice these skills outside the group.

After the education phase, the group discusses the rehearsal phase. The last few discussions should have helped the group members identify some of the stimuli that cause them to emit their negative behaviors. The members should be encouraged to determine coping statements and to practice. The homework for this meeting is to go into the application phase and use these statements in their daily life.

Session 5

This discussion should begin by reminding the group this is the next to last group meeting. The members and the discussion should focus on how the different techniques taught during the group have helped them in their lives. The group should discuss how the techniques learned can help with their issues of power and control. The group should be very reinforcing to the members at this point, and the discussion and sharing should focus on that aspect.

The second portion should be a discussion of how the members applied their new cognitive restructuring technique. Attention should be given to both alternative strategies all the members could benefit from and to each member for progress they have made. The therapist's main role during this discussion is to reinforce all the members of the group for trying this new technique and for sticking with the group. It should be pointed out that mastery of this technique will only come with practice, and these thought processes will only become inherent by overtly practicing them in day to day situations.

Session 6

The discussion for this meeting should be totally devoted to the termination of the group. The members will be asked to identify sources of support, reflect on whether they feel the time is right for long-term group therapy, if they feel they need individual follow-up appointments with the therapist, and, most importantly, how the group has affected them. The members should also discuss their newly developed skills and how to practice them. Consideration should be given to every group member so none will be left out.

Although this short-term therapy group has many advantages, long-term treatment is essential for this population. Singer (1989) writes "short-term treatment is not adequate in helping them to identify the emotional and behavioral symptoms, recognize the negative messages from the past, and develop ways to alter the feelings and behavior that work against them" (p.

470). This type of therapy can be affective by building trust between members, decreasing feelings of isolation common among incest survivors, and demonstrating strategies to actively gain control. In essence, this therapy is important to help survivors recognize and change dysfunctional feelings, beliefs, and behaviors.

References

Josephson, G., & Fong-Beyette, M. (1987). Factors assisting female clients' disclosure of incest during counseling. Journal of Counseling and Development, 65, 475-478.

Joy, S. (1987). Retrospective presentations of incest: Treatment strategies for use with adult women. Journal of Counseling and Development, 65, 317-319.

Morrow, K., & Sorell, G. (1989). Factors affecting self-esteem, depression, and negative behaviors in sexually abused female adolescents. Journal of Marriage and the Family, 51, 677-686.

Singer, K. (1989). Group work with men who experienced incest in childhood. American Journal of Orthopsychiatry, 59(3), 468-471.

Sparks, A., & Goldberg, J. (1994). A current perspective on short-term groups for incest survivors. Women & Therapy, 15(2), 135-147.

Thorpe, G., & Olson, S. (1990). Behavior Therapy. Needham Heights, MA: Allyn and Bacon.

Group Therapy for Rape Survivors: A Combination of Person-Centered and Cognitive-Behavioral Techniques

Michael V. Waggle
St. Mary's University

Abstract

Rape is a crime that is increasing at an alarming rate. The psychological effects of rape generally leave the survivor in a state of confusion. A rape survivor often feels she has lost control of her life, is generally embarrassed, and has trouble trusting others. There is a need for specialized psychological services for rape survivors. With the increasing costs of medical services, group psychotherapy is an economical and efficient alternative to prolonged medical care. This paper is intended to present a theoretical outline for a rape support group. The sessions are divided into person-centered discussions and cognitive-behavioral techniques designed to allow the survivor to regain some control over her situation.

The statistics documenting this crime are appalling. FBI statistics show that in 1979 there were "75,989 official reports of rape" (cited in Kilpatrick, Resick, & Veronen, 1981, p. 106). In a study of college students, Mills and Granoff (1992) state "approximately one in every five female college students will be victims of sexual assault before they graduate" (p.504). The most disturbing statistic is cited by Resnick and Newton (1992) "results of epidemiological studies of nonclinical populations indicate that between one-fourth and one-half of women surveyed had experienced some form of sexual assault" (p.100). The long term effects of rape trauma include major depression, anger, hypervigilance to danger, and sexual dysfunction (Frank et al., 1988). The survivor feels a loss of control over her life.

Sutherland and Scherl (cited in Kilpatrick et al., 1981 and Yassen & Glass, 1984) developed a three stage model to describe the recovery process of the rape survivor. The first stage, acute reaction, is characterized by disorganized behavior and a complete disruption of post-rape life. This stage normally lasts between 2 to 4 months. Frank , Anderson, Stewart, Dancu, Hughes and West (1988) indicate that the immediate reactions to rape are generally somatic. Some of these somatic responses include back pain, weight fluctuations, stomach pains, and menstrual irregularities (Kimerling and Calhoun, 1994). This could explain why most rape survivors seek medical attention rather than seeking counseling services.

The intermediate stage of Sutherland and Scherl's model involves denial and an outward adjustment. Therapy termination is common in this stage since the client, on a superficial level, seems to have recovered. Nightmares, phobias, and sexual dissatisfaction are common. Between 3 months to 1 year post-rape, the individual's symptoms are focused on rape-specific anxieties. The information processing theory describes these anxieties as information stored in fear networks (Resick and Schnicke, 1992). This theory explains that memories of the trauma are stored in "fear networks that consist of stimuli, responses, and the meanings of the stimulus and response elements" (Resick and Schnicke, 1992, p.748). The stimuli stored in these fear networks can range from the location of the rape, time of day, and/or the general appearance of the rapist. This fear network develops a fear schema in which most rape-related clues elicit fear and anxiety. The survivor is then predisposed to attend to evidence consistent with the predisposed fear schema and ignores those clues that are inconsistent with the schema.

Finally the survivor transitions to an integration and resolution stage. As the individual works to integrate the experiences of the past, a rape-clued stimuli may greatly decrease the survivor's functioning. The client usually does not return to therapy at this time, as she fears a sense of failure.

Some symptoms are prevalent throughout all these stages. The survivor normally shows signs of extreme guilt, impaired trust, and fear (Mills and Granoff, 1992). The survivor's fear is usually generalized into a fear of being alone, fear of the dark, or some phobic reactions. The survivor also exhibits an increase of operant avoidance behaviors which include never being alone, sleeping with lights on, and avoiding men she doesn't know.

Resnick and Newton (1992) show that rape trauma generally increases many interpersonal issues for the survivor. Social support is vital in rape cases, because this support has been shown to act as a buffer for some of the psychological ramifications of rape trauma (Kimerling and Calhoun, 1994). Many times the family, although wanting to help the survivor, feels uncomfortable discussing the matter. This silencing norm may increase the survivor's feelings of guilt, shame, and isolation. Group therapy can be a very profitable avenue for examining these issues. Resnick and Newton (1992) have found that the opportunity to share the rape experiences and consequences are very valuable for the survivors. The group experience can be used to decrease many of the rape-phobic reactions, generalized anxiety, and the survivor's central need for approval. The group experience tends to decrease feelings of isolation and alienation common in most members. Roth, Dye, and Lebowitz (1988) found that "For the members of the group, the experience of feeling overwhelmed and then managing or profiting from it, and seeing other people also survive these periods of intense affect, was encouraging and empowering" (p.85).

Group work is not advisable for those clients seeking treatment immediately after the rape. According to Yassen and Glass (1984) the client in the acute stage is more concerned with meeting daily routine needs than interpersonal needs. If groups contain women in the acute stage of rape trauma, much of the group process is likely to be geared toward taking care of these women. Women in the acute stage may even be further traumatized by seeing women still needing help months or even years post-rape. For these concerns it is advisable that groups only contain members well into the outward adjustment and denial stage. Immediately following rape-trauma the survivor should receive individual counseling and then, at approximately the 5- to 6-month point, be referred for group therapy. This shift in therapeutic techniques may keep the survivor in therapy longer and; therefore, lessen the anxiety felt throughout the integration resolution stage.

Several other factors may influence the acceptance of members for group therapy. Survivors of group rapes, rapes by more than one perpetrator, may be included in groups with survivors of individual rapes. Gidycz and Koss (1990) found that the survivors of both individual and group rapes have similar symptomologies on the Trait Anxiety Inventory and the Beck Depression Inventory. Incest survivors, however, should not be included in groups with rape survivors. Sharma and Cheatham (1986) found that the primary symptom of rape survivors is fear, but Resnick and Jordan (1988) found the primary symptom of incest survivors is a loss of trust and identity confusion.

Yassen and Glass (1984) discuss that location and meeting time is particularly important for this type of group. The group should not meet in a place that may resurface rape-anxiety cues. Hospitals, isolated buildings, and areas that may be perceived as dangerous are particularly poor choices since they may elicit rape-cued memories. The group should meet in a location and at a time to allow the members to come to and leave the group during the day. The sessions should last for 9 weeks. Group sessions should be held once a week for 2 hours. The first 50 minutes should entail a person-centered type discussion of prearranged topics. This first portion of the meeting should allow the members to discover the deep feelings elicited by the rape-trauma. The next 50 minutes should be used to teach the members cognitive-behavioral techniques to manage these emotions.

The goals for the group can be short-term and long-term. The short-term goal for therapy should be to develop trust among the group members. By developing trust the members should feel safe to acknowledge the rape-trauma and disclose the consequences of the rape. The long-term goal for therapy is control. The cognitive-behavioral aspect of this treatment should be specifically designed to accommodate this goal. These techniques should give the survivor something active to do during therapy. By giving the individual control over her therapy and over her emotions, she should gain some control over her symptoms and over her life. The cognitive-behavioral techniques used in this theoretical group are exposure-type therapy with an emphasis on coping skills (Resnick and Newton, 1992) and an adaptation of Meichenbaum's stress inoculation training (Resick and Jordan, 1988).

All techniques should be applied to a non-rape stimuli and then to the rape stimuli. The therapy should be divided into three phases. The first phase, sessions 1 and 2, is the Education phase. Clients first learn some of the phases and symptoms associated with rape and then learn the basics of cognitive-behavioral therapy. The second phase of therapy, sessions 3-5, is the Skill Building phase. Finally the therapy moves to an Application phase, sessions 6-9.

Session 1 (Education phase)

The first session should begin with some evaluative measures. The clients should be given the Beck Depression Inventory to measure depression, the SCL-90-R to check for somatization and distress, and the FIRO-B to examine interpersonal relationships. These three measures were chosen to evaluate the client's functioning over a wide span of modalities. The first discussion should focus on the phases and symptoms of rape survivors and then on matters of trust. Trust building exercises should be guided toward the clients discussing the characteristics of the rape-trauma. These commonalties should be used to build trust throughout the group.

The first behavioral strategy to be employed should be controlled breathing exercises. Clients should be taught to calm themselves by taking deep controlled breaths. As the clients imagine a non-rape anxiety provoking situation they should take deep diaphragmatic breaths. Each breath should take approximately 3 to 4 seconds for inhalation and exhalation (Resnick and Newton, 1992). As the clients calm down, attention should be paid to the control the clients have over their emotional response. By controlling their emotions the clients should learn that there is a difference between thoughts and emotions and that they can control each.

Session 2

This guided discussion should focus on the schemata of self and anger by dealing with the ideas of self-esteem and emotions. Clients should be asked for a schemata of self before and after the rape. They should be asked how the rape has changed the way they deal with situations. The second behavioral technique employed should be a modification of Jacobsonian muscle relaxation. The clients should focus on a pleasant scene and then stress specific muscle groups. The clients should be instructed not to let thoughts ramble since most rape survivors thoughts will focus on anxiety producing events (Frank et al., 1988). Through this exercise the clients should learn that they have control over their physical responses. Through the controlled breathing and muscle relaxation techniques, clients should be taught the contrast between anxiety and relaxation.

Session 3 (Skill Building phase)

This discussion should be focused on the type of stimuli that tend to produce anxiety. Particular attention should be paid to the similarities of the group, and how each member can counteract these fears. If the fears are irrational, the focus of the group should be on the irrationality of some linked fears. The second portion of the therapy should focus on how to use some of the previously learned techniques. The client should be exposed to the three anxiety response channels: (1) physical reactions (2) thoughts of threat and (3) survival behaviors (Resnick and Newton, 1992). The client should be taught to focus on which channel is being activated and then should be instructed to use one of the previous techniques to lessen the

anxiety. It is important to emphasize that the client should gain control of the situation by staying in it. At this time it is also important to stress that whatever the clients did to survive the rape-trauma was correct. Time should not be spent wondering "what if."

Session 4

This discussion should be focused on safety concerns. The clients should be encouraged to describe the situations in which they do not feel safe and how the rape changed the way they view the world. Common myths like "rape only happens to bad girls" and "I will never be safe" should be explored. Attention should be focused on ways in which the clients may protect themselves, i.e. having door locks checked, traveling with friends, and using mace.

This session should be used to teach the clients covert rehearsal and guided self-dialogue techniques. Covert rehearsal requires that the clients imagine themselves in a stressful non-rape situation. The clients should mentally rehearse coping with the situation by using some of the previously learned situations. The guided self-dialogue technique allows the clients to assess and replace irrational and dysfunctional cognitions with cognitions that promote functionality and coping skills (Resnick and Newton, 1992). Resick and Jordan (1988) describe this technique by having clients first learn that self-defeating private dialogues generally increase anxiety. By defeating this habit, clients learn they can reduce their anxieties. The first step in this technique should be to define the problem. The problem should then be broken down into several small steps and all of the client's attention should be focused at doing each tasks. Finally the clients should be instructed to be self-reinforcing.

Session 5

In this session the clients should be asked to describe the ways they feel they have lost power and control in their lives, and how they can use some previously learned techniques to gain more control. Their perceptions of their experience and their individual roles in these perceptions should be paramount to this discussion. It should be reinforced to the clients that they were not responsible for the rape-trauma and that, no matter the individual situation, the perpetrator had no right to commit the crime.

Now that the clients should have learned the techniques involved with mentally working through situations, they should transition to overtly working through situations by role playing. Resick and Jordan (1988) break this technique into the following steps: assessing the reality (danger) of the situation, controlling negative thoughts and self-statements, acknowledging, using, and possibly relabelling the experienced arousal, preparing for the feared cue, using the coping skills to confront the feared cue, and reinforcing themselves for coping. (p.104) The therapist's role should be to reinforce behaviors, make suggestions, and model the role-playing behaviors. The homework assignment is to actually act out an anxiety-producing situation.

Session 6 (Application phase)

This entire session should be devoted to developing the fear hierarchies that were started in session 3. The clients should break the fear-producing situations into manageable units much like what should have been accomplished in the guided self-dialogue technique. It is important that the client's fear hierarchy have several dimensions. A typical fear hierarchy should progress in number, sex, familiarity, and time of day (Resick and Jordan, 1988).

Session 7

The discussion in this session should focus primarily on issues of intimacy and guilt. The clients should be encouraged to examine their relationships to see if they have changed since the

rape-trauma. If they have, the clients should be encouraged to see how they can use the previous techniques to work through their issues. The session should be used to teach the clients how to solve problems creatively. Many rape survivors feel immobilized to make any decision. Since rape survivors may be quick to disregard suggestions, the therapist must create a brainstorming norm where the clients allow all possible alternatives. Both short- and long-term consequences should be evaluated. The group members may be reluctant to follow through on some alternatives, but they should be encouraged to use some of the previous techniques to accomplish tasks.

Session 8

The three measures given during session 1 should be given again to measure the individual's psychological stress at this point in therapy. This guided discussion should be left open to the group. Any topic that a member wants to bring to the group should be addressed. This session should also be used to discuss closure of the group and to bring all aspects of the program together in the stress inoculation process. The clients should be shown how they can initially identify a fear condition, identify the cue for the fear, and then control their fears. Resick and Jordan (1988) describe Meichenbaum's stress inoculation procedures as follows: assessing the actual probability of the negative event happening again, managing the overwhelming avoidance behavior, controlling self-criticism and self-devaluation, engaging in the feared behavior, reinforcing self for attempting the behavior and for following the protocol. (P. 109)

Session 9

This entire session should be used to discuss the termination of the group, review the program, and discuss any changes in the scores on the psychological measures. The clients should be instructed that these skills must be practiced regularly. Special care should be taken to have inputs from all members on how the program has worked for them and things on which they need to work.

Conclusion

This therapy should be effective for rape survivors because it addresses all aspects of their lives. The discussions at the beginning of the sessions are geared to build trust between members, bring members in touch with their feelings, and decrease the sense of isolation commonly found with rape survivors. The cognitive-behavioral portion of the theoretical group should help survivors to be active participants in their recovery and; therefore, help them to decrease the common anxiety and avoidance behaviors associated with post-rape functioning.

References

Frank, E., Anderson, B., Stewart, B. D., Dancu, C., Hughes, C., & West, D. (1988). Efficacy of cognitive behavior therapy and systematic desensitization in the treatment of rape trauma. Behavior Therapy, 19(3), 403-420.

Gidycz, C. A., & Koss, M. P. (1990). A comparison of group and individual sexual assault victims. Psychology of Women Quarterly, 14(3), 325-342.

Kilpatrick, D. G., Resick, P. A., & Veronen, L. J. (1981). Effects of a rape experience: A longitudinal study. Journal of Social Issues, 37(4), 105-122.

Kimerling, R., & Calhoun, K. S. (1994). Somatic symptoms, social support, and treatment seeking among sexual assault victims. Journal of Consulting & Clinical Psychology, 62(2), 333-340.

Mills, C. S., & Granoff, B. J. (1992). Date and acquaintance rape among a sample of college students. Social Work, 37(6), 504-509.

Resick, P. A., & Jordan, C. G. (1988). Group stress inoculation training for victims of sexual assault: A therapist manual. In P. A. Keller & S. T. Heyman (Eds.), Innovations in Clinical Practice: A Source Book, Vol. 7 (pp. 99-111). Sarasota, FL: Professional Resource Exchange.

Resick, P. A., & Schnicke, M. K. (1992). Cognitive processing therapy for sexual assault victims. Journal of Consulting & Clinical Psychology, 60(5), 748-756.

Resnick, H. S., & Newton, T. (1992). Assessment and treatment of post-traumatic stress disorder in adult survivors of sexual assault. In D. W. Foy (Ed.), Treating PTSD: Cognitive-behavioral strategies, Treatment manuals for practitioners (pp. 99-126). New York: Guilford Press.

Roth, S., Dye, E., & Lebowitz, L. (1988). Group therapy for sexual-assault victims. Psychotherapy, 25(1), 82-93.

Sharma, A., & Cheatham, H. E. (1986). A women's center support group for sexual assault victims. Journal of Counseling & Development, 64(8), 525-527.

Yassen, J., & Glass, L. (1984). Sexual assault survivors groups: A feminist practice perspective. Social Work, 29(3), 252-257.

Correlates of Course and Faculty Perceived Effectiveness

Michael J. Benson & David B. Porter, DPhil
Department of Behavioral Sciences and Leadership
United States Air Force Academy

One hundred fifteen faculty members at the US Air Force Academy completed demographic and personality questionnaires and submitted these together with their end-of-semester student course critiques. These inputs were examined to determine the strongest predictors of positive student ratings. Three variables were found to have particularly strong relations with the effectiveness ratings faculty members received from their students: years of teaching at the Academy, the teacher's relative emphasis on students' attitudes rather than their knowledge and the teacher's Myers-Briggs "temperament". This paper, describes the relevant findings, and discusses the implications for further research.

The quest for quality seems to have extended to every facet of American enterprise in the past decade (Aguayo, 1990). Even in the ivory towers of academe, recognition of the importance of understanding critical processes and satisfying "customers" is becoming increasingly common place (Marchese, 1992; Porter, 1991). Although there is still considerable debate concerning who the "real" customers of education are, the use (and occasional abuse) of student course critiques has continued to grow. Even if students are not the ultimate customer of education, they are clearly "stakeholders" and perhaps even "willing workers" in the process of education. As such, their attitudes, as measured by their responses to items concerning the effectiveness and quality of their instructors and curriculum are important. The importance of such data does not rest on its objective validity alone (although there is now considerable evidence that critique ratings do show moderate correlations with objective measures of student learning (Cashin, 1990; Porter, 1992)). The real importance of critique data are the consequences it is likely to have on students' future choices and decisions. In fact, student perceptions of courses tend to become more pronounced over time. As Joe Petit (cited in Cisneros, 1992) found from visiting various alumni reunions, courses that had been perceived as "very good" by students were remembered as having been "great" by graduates 20 years later. Conversely, bad courses seemed to grow worse with the passage of time. Most of the average courses, unfortunately, were forgotten altogether.

The United States Air Force Academy initiated a program of comprehensive course critiques several years ago (Porter, 1988). However, concerns of both individual faculty members and administrators largely prevented meaningful research at the individual faculty member level across disciplines. The emergence of the Educational Outcomes Assessment Working Group in 1994 reflected a renewed commitment to better understanding educational processes. One of the first projects undertaken by this group was a survey of all faculty members concerning their attitudes and teaching practices. To encourage wide participation, complete anonymity was promised to participants. Even with such anonymity, several senior faculty members were uncomfortable with the inclusion of items relating to the personality of individual faculty members. This initial survey was moderately successful. Garnering a 57 percent participation rate, its results reflected marked differences in attitudes, perspectives and practices across

academic departments and divisions. This sample was sufficiently large to allow meaningful analysis of departmental faculty characteristics as predictors of departmental average critique ratings.

The present study was developed as a supplement to the larger study. By requesting individual voluntary participation, this study sought to examine the relations between several different individual difference measures and the individual ratings instructors received from their students. Three types of individual difference measures were used as independent variables: demographic, personality (the Myers-Briggs Personality Type Indicator, (MBTI)) and epistemological. Gender, race, degree level, academic and military rank and years at the Academy were among the demographic data collected. Individual dimensions of the MBTI (i.e., I-E, S-N, T-F, & J-P) as well as relative affinity for the four temperamental dispositions (i.e., SP, SJ, NT, & NF) served as individual difference predictors (Myers & McCaulley, 1985). A full explanation of the Myers Briggs Type Indicator Scales and the Kiersey Bates Temperaments is beyond the scope of this paper; interested readers are directed to Myers and McCaulley (1985) and Kiersey Bates (1978) as excellent references. Separate composite scores were used as dependent variables; one was a measure of the instructors perceived effectiveness and the other reflected students perceptions of course effectiveness.

Methods

Notice of this experiment was sent to all faculty over the local area network. The invitation to participate contained a brief synopsis of the study including the instruments and procedures to be used. Participating faculty would be asked to complete a three page questionnaire containing general demographic descriptions as well as locally developed indicators of the dimensions and temperaments of the Myers Briggs Personality Type Indicator. Upon receipt of a faculty members' willingness to participate, the three page questionnaire was sent to them with instructions to complete it as soon as possible and retain it until individual course critique information became available approximately 3 weeks later. A copy of the critique information was then to be returned with the previously completed questionnaire. Data were compiled and analyzed using standard Statistical Programs for the Social Sciences (SPSS). An initial summary of demographics and personality averages and variation along with individual scores were returned to participants approximately 2 months after their submission of forms. This report was also distributed to participants several months later.

Results

The 115 faculty members who participated represent approximately 23% of the faculty. Initial concerns that only those who received favorable critique ratings were allayed by the discovery that sample averages on all critique items were very close to but slightly below overall faculty averages. Slightly more women (20.3%) but slightly fewer racial minorities (5.9%) participated than would have been expected (both groups comprise approximately 10 percent of the faculty). The faculty is divided into four academic divisions of approximately equal size; however, participation by academic division was somewhat uneven. The 41 responses from the Basic Sciences Division made up 35.7% of the sample while the 11 responses from the

Humanities Division were only 9.4% of the sample. The other two divisions showed approximately representative participation rates: Social Sciences' 33 responses were 28.0% and Engineering's 30 responses were 25.4% of the sample. Results will be presented and discussed in two sections; one for each of the individual differences examined: demographic and personality variables.

Both Instructor Perceived Effectiveness and Course Perceived Effectiveness were the average of student responses to five separate items of the 26-item US Air Force Academy Course Critique. The five items included in Instructor Effectiveness were: Instructor's ability to stimulate student interest; Instructor's ability to provide clear, well-organized instruction; Instructor's concern for student learning; Instructor's enthusiasm and Instructor's effectiveness in facilitating learning. The five items included in the Course Effectiveness were: The degree to which the course met its stated objectives; the intellectual challenge and encouragement of independent thought; Evaluative and grading techniques (tests, papers, & projects); Quality and usefulness of course; and The course as a whole. Initial analysis revealed that the correlation between these two variables was .838. Although this relationship means that 70% of the variance in these two criteria was common, the 30% of the variance which was independent might provide useful indications concerning causality and validity.

Demographic:

The lack of experience of the Academy faculty was apparent in measures of total years teaching, years teaching at the Academy, and military and academic rank. Although the mean number of years of college teaching was 5.1 (SD = 5.4), the median was 3 years and the modal response was 1 year. Similarly, the mean Academy teaching experience was 3.45 (SD = 4.0), but the median response was 2 years and the mode was again 1 year. Table 1 reflects this strong positive skew:

Table 1
Experience teaching college and at USAFA (% of sample)

Total Years:	0-2	3-5	6-8	9-11	12-14	15-17	18-20	20+
College Teaching	43.5	21.7	20.9	.9	7.0	.9	2.6	2.6
USAFA Teaching	54.7	21.2	14.4	5.8	2.4	0	0	0

Assistant Professors (45.8%) made up the largest portion of the sample with Instructors (29.7%), and Associate Professors (15.3%) accounting for most of the rest of the sample. Six full professors (5.1% of the sample) also participated. Thirteen civilian faculty members (11%) participated in the study. The 89% in the military approximated the rank distribution on the faculty: 42% Captains, 34% Majors, 19% Lieutenant Colonels and 4% Colonels. Half the sample had earned doctorate degrees and half had masters degrees. Approximately 25 percent of the sample had earned aeronautical ratings as pilots (12.8 percent) or navigators (12.8 percent) and

58.3% had earned Bachelors degrees from the Air Force Academy. These last two factors are both assumed by some to be advantages in garnering high critique ratings because of the greater identification of students with individuals who match their aspirations (i.e., to graduate from the Academy and fly).

Table 2 contains the Pearson correlations between different demographic predictors and instructor and course critique Perceived Effectiveness Scores. For Gender, males were scored as "1" and females "0"; for Rating, PhD and USAFA Grad a "1" was used to reflect status and a "0" for its absence.

Table 2.
Correlations between Demographic Factors and Perceived Course and Instructor Effectiveness

Demographic Factors:	Height	Gender	Aero Rating	Degree (PhD)	USAFA Grad	Mil Rank	Acad Rank	Yrs@ USAFA	Yrs Col Tching
Course PES	.15	.20*	.08	.16	.07	.26**	.38**	.38**	.30**
Instructor PES	.18	.21*	.09	.13	.02	.26**	.35**	.38**	.29**

* p<.05 **p<.01

Several potential predictors had nonsignificant correlations with students' perceived effectiveness of courses or instructors. Instructor's height, aeronautical rating, academic degree and alumnus status all failed to show significant relationships to either criteria. Although the relations between gender and perceived effectiveness were significant at the .05 level, the size of these correlations suggests that their effects are relatively small (i.e., they explain only 4% of the variance). In contrast, academic rank and years of experience at the Academy showed moderate correlations which were significant at the .01 level and explained approximately 15% of the variance in criteria. The slightly lower correlation between Years of College teaching and perceived effectiveness suggests that while teaching experience at other colleges and universities may be useful, it is a less potent predictor than experience at the Academy.

MBTI (Personality Variables):

Average scores of participants on the four underlying dimensions and four temperaments were divided by the standard deviation of the sample distribution of those scores. The resultant standard score represented displacement from a neutral score. In the case of the four underlying dimensions, preference for the first letter in the pair (i.e., E, S, T or J) would be reflected by a negative standard score. Positive standard scores reflect average preferences for the second letter in the pair (i.e., I, N, F, or P). For the temperaments, scores represent relative affinity. Higher scores reflect stronger general endorsement of the interactional style associated with the particular temperament. Negative scores reflect relative dislike for or discomfort with the respective style. Table 3 provides the correlations with the perceived effectiveness scores for both courses and instructors in the two rows below the standard scores.

The sample's standard scores for the basic dimensions of the MBTI suggest that, as a group, the USAFA faculty are nearly equally divided between Extroversion and Introversion, somewhat more oriented toward Sensing than iNtuition, strongly prefer quantitative criteria (Thinking) to qualitative ones (Feeling) in decision making and showed a moderate preference for making decisions sooner (Judging) rather than later (Perceiving). Although the average faculty score represented a moderate preference for Sensing over iNtuition, the correlations with both course and instructor perceived effectiveness scores suggest that those faculty members who had a greater preference for patterns rather than particulars (i.e., N rather than S), received significantly higher ratings from students. This relationship explained about 7.5% of the total variance in critique ratings. Extroversion-Introversion did not show a substantial relation to rated effectiveness, nor did Thinking-Feeling. Instructors preference for Perceiving (i.e., deferring decisions to gather more information) appeared to be somewhat positively related to students' rating of their effectiveness but this relationship accounted for less than 4% of the variance.

Table 3
Standard Scores and Correlations between MBTI Dimensions and Temperaments and Perceived Course and Instructor Effectiveness Ratings

Dimensions:	E-I	S-N	T-F	J-P	Temps: SP	SJ	NT	NF
Std Score:	-.05	-.39	-.83	-.23	-.83	.19	1.27	-.50
r w/CrsPES	.16	.27**	.14	.14	-.19*	-.29**	.29**	.26**
r w/InstPES	.07	.28**	.14	.18*	-.17	-.30**	.27**	.27**

* p<.05 **p<.01

Although temperaments are constructed from personal preferences along the underlying dimensions, they reflect a higher level of organization and thus may vary from the dimensional scores. Temperament standard score averages suggest that the traditional academic NT perspective received the strongest endorsement from the faculty, while the action-oriented SP temperament was the most strongly rejected. The stable, tradition of the SJ received slight endorsement and the idealism of the NF temperament, mild rejection. Despite the fact that faculty appeared to slightly prefer Sensing to iNtuition as an underlying perceptual style, they strongly endorsed the NT temperament. Examining the percentage of the sample who were categorized in each of the four temperaments reiterates the distinctions suggested by the standard scores. Inquisitive NTs made up 60.2% of the sample and traditional SJs accounted for another 23.7%. Idealistic NFs were 11.0 % of the sample but energetic SPs accounted for only 5.1%. The instrument used to gather these data has also been administered to cadets as part of several different courses over the last decade. Although young adults characteristically show a great deal of individual personality volatility, the average proportion scoring in each of the four temperaments on the same instrument used in this study has remained relatively equal and constant (i.e., 25% NTs, 25% SJs, 25% NFs and 25% SPs). The relations between temperament and critiques are consistent with the correlations found on the underlying MBTI dimensions: affinity for the abstract temperaments (either NT or NF) were positively correlated

with course and instructor critiques. Affinity for the stable, traditional and well-organized SJ temperament was negatively correlated with critique ratings. Affinity for the SP temperament was also slightly negatively correlated with critique scores.

Together, results from both these analyses suggest that individual MBTI scores show several moderate correlations with perceived teaching and course effectiveness. However, the data suggest that the most "obvious" personality advantage, that of being an extrovert, was of no consequence in getting higher ratings. In contrast, the preference for iNtuition, or a desire to put things together and to present broader themes and larger patterns, was consistently associated with higher ratings from students. Whether faculty relied on logical, quantitative criteria or more value-oriented, qualitative standards did not seem to make much difference to students. There was also some indication that students react negatively to what they may perceive as premature closure (J-ness) or too frequent attempts to assert authority and control in the classroom (SJ-ness).

Discussion

This study complements the other assessment activities undertaken by the Educational Outcomes Assessment Working Group. It's unique contribution is the exploration of correlations between individual differences variables and course and instructor critique ratings. Correlation is not causation and educational processes are too complex to yield entirely to such simple analysis. However, this study does provide information relevant to a number of educational practices and processes; consequently, it is a step in the right direction.

Support was found for the notion that individual faculty members increased their perceived effectiveness with increased tenure. Interestingly, a similar study conducted in the first year of adoption of the initial comprehensive course critique program (Porter, 1992) showed no significant relationship between military rank and rated classroom effectiveness. This might be an example of the influence of the adoption of a metric such as course critiques on institutional processes such as faculty development and selection for continuation. The fact that being an Academy alumnus, rated officer or possessing a doctorate degree does not significantly increase an instructors' perceived effectiveness is also potentially useful policy information. The news concerning the relation between gender and ratings is both bad and good. The bad news is that such differences persist; the good news is that the differences are relatively small (i.e., associated with only about 4% of the variance in ratings). In fact, the correlation between height and perceived effectiveness was very close to the correlation between perceived effectiveness and gender. Additionally, the correlation between gender and height was also very strong ($r=.68$, $p<.01$). It is at least possible, the slightly lower ratings for women faculty members are as much a reflection of an unconscious bias based on height rather than a general lack of respect based on their gender.

Examinations of correlations between personality measures and rated effectiveness approached but did not exceed the common ceiling of .33 or 10% of the variance. Perhaps the lesson to be learned from these results is not to try and identify or select individual ideal teachers but to recognize the advantages of recruiting faculty with more diverse styles and perspectives.

There is a significant danger of like-minded individuals convincing themselves that their personal preferences are normative rather than merely descriptive. The attempt to impose these stylistic preferences on others, especially those with other preferences is likely to create resistance and adversity and may impede learning. The relatively great diversity of temperaments of Academy cadets suggests that classrooms which include a variety of different learning activities and alternative ways for students to demonstrate competence may be more favorably evaluated by students.

Implications for Future Research

These initial findings serve the purpose of organizing and describing the data. The next steps taken with these data will be the most important. One potential path exists in the findings of Carol Dweck (Dweck & Leggett, 1988) who suggested that the way students think about intelligence largely determines how they behave in the classroom and thus what they learn. Students who view intelligence as a skill that can be developed through interaction and exploration, are much more likely to actively participate in activities and discussions, even if they don't already know "the right answer". Not too surprisingly, students who believe intelligence is learnable often do learn more from their courses because of the quantity and quality of their class participation. It is probable for instructors to harbor similar views; therefore, it may be beneficial to look into the aspects of instructors' epistemological paradigm, and evaluate the effects on perceived effectiveness.

Another potential area of analysis rests in the aspect of combined effects and divisional effects. Ultimately, the goal is to improve the institution. We may be able to learn valuable lessons from some of our own faculty. For this reason, studies focused on variance in perspectives within departments and even within courses are likely to reveal additional information, insight and opportunities for improvement. The Air Force Academy has a mission of inspiring air and space leaders with vision for tomorrow, hopefully the institution will take advantage of the opportunity to learn about itself; consequently improving the process of learning.

References

Aguayo, R. (1990). Dr. Deming: The American who taught the Japanese about Quality. New York: Carol.

Cashin, W.E. (1990). Student ratings of teaching: A summary of the research. Management Newsletter 4(1). pp. 2-7.

Cisneros, H. (1992, June) Higher Education: the Cutting Edge. Bulletin of the American Association of Higher Education.

Dweck, C.S. & Leggett, E.L. (1988). A Social-Cognitive Approach to Motivation and Personality, Psychological Review, Vol 95. pp. 256-273.

Kiersey, D. & Bates, M. (1978). Please Understand Me: Character and Temperament Types (3rd ed.). Del Mar, CA: Prometheus Nemesis Books

Marchese, T. (1992). AAHE and TQM (...make that "CQI"). American Association of Higher education Bulletin, 45(3).

Myers, I.B. and McCaulley. (1985). Manual: A guide to the development and use of the Myers Briggs Type Indicator. Palo Alto, CA: Consulting Psychologists Press.

Porter, D.B. (1988). Course Critiques: What Students can tell us about educational efficacy. Proceedings of the human Factors Society - 32nd Annual Meeting. Santa Monica, CA: Human Factors Society.

Porter, D.B. (1991). Total Quality Education. In Lam K.D., Watson, F.D., & Schmidt, S.R., (Ed.) Total Quality: A textbook of strategic quality leadership and planning. Colorado Springs, CO: Air Academy Press.

Porter, D.B. (1992). Student Course critiques: A case study in total quality education. Proceedings: Psychology in the Department of Defense Thirteenth Symposium. USAF Academy, CO: Department of Behavioral Sciences and Leadership.

Cadet Attitudes about Group Work

Justin A. Hansen & David B. Porter
United Stated Air Force Academy

This study examined cadet attitudes concerning group work in academic courses. One hundred six junior and senior cadets were asked how the existing level of group work at the Air Force Academy should be adjusted to improve educational effectiveness. Three independent predictors of cadet attitudes were found: 1) grade point average was inversely related to cadets' attitudes concerning academic group work; 2) females disliked group work more than males; and 3) seniors disliked group work more than juniors. The combined linear effects of these three factors explained 29 percent of the variance in cadet attitudes toward group work.

Group work is used extensively at the Air Force Academy, and every cadet has a great deal of exposure to it, regardless of academic major. There are several types of group work assigned to cadets both in and out of class. The type most commonly used is out of class group work. There is an extensive literature on Cooperative Learning, Group Discussion, Problem Based Learning, and group interaction relevant to this study (Bruffee, 1987; Slavin, 1987; Lyman & Foyle, 1990; Savoie & Hughes, 1994; Porter, 1989).

In "Cooperative Learning: Student Teams", Slavin (1987) discusses several studies that show students learn more by working cooperatively than competitively. When students' grades are based not upon a result or product, but instead on the entire group's understanding of material, students tend to perform better and get more out of the experience (Kohn, 1986). This is attributed to students' willingness to help each other learn and understand the material. In contrast, students in competetively oriented classrooms are much less likely to help each other. Other studies have shown there are certain essential task characteristics for group activities to increase students' learning or understanding(Stahl, 1994).

Other research shows that different types of group activities are better for different subjects. The technique that helps students learn vocabulary words in an English class may not be the same technique that helps students derive formulas or complete a lab for mathematics or physics. Research has shown that group work is also slightly more beneficial to minorities, including women and certain ethnic groups (Slavin, 1987). Attitudes and expectancies are potent predictors of performance. For this reason, understanding what influences students' attitudes toward groups is important because attitudes often predict group success. This study was an attempt to discover the determinants of cadet attitudes towards the use of academic groups at the Air Force Academy. There are indications that student teams build social skills in younger students as well, such as cooperation and group cohesiveness (Kohn, 1986). These are important because cadets need to develop a positive orientation toward working with others to achieve success as Air Force officers.

Methods

The participants in this study were junior and senior cadets, ranging in ages from 19 to 22. There were 106 participants in all, including 86 males and 20 females. (A slightly higher representation than the 13% base rate.) The sample included 45 juniors and 61 seniors. The participants were chosen at random from several different cadet squadrons. The instrument used was a simple survey composed of four questions, as well as demographic information. Demographic information included class, gender, academic major, whether the subject was an intercollegiate athlete, and the subject's grade point average(GPA). Academic majors were divided into two categories, "technical" and "non-technical". Cadets are all required to take a very extensive core curriculum from four different academic divisions (Humanities, Engineering, Basic Sciences, and Social Sciences) each of which contribute roughly 6-8 classes to the "core" curriculum. All academic majors sponsored by Humanities or Social Sciences were considered "non-technical", while those sponsored by either Engineering or Basic Sciences were considered "technical". The questions on the survey were prefaced by a brief introduction and explanation of the survey's purpose.

The criterion question from the survey asked respondents how the Academy could increase its educational effectiveness with respect to the use of out-of-class group work. Subjects were given five response choices ranging from "greatly increasing the use of out of class group work", to keeping it the same, back to "greatly decreasing the use of out of class group work". The potential effects of each of the five demographic variables (class year, gender, major, athlete status, and grade point average) were examined using regression analysis, as well as interactions and curvilinear relations between them (Cohen & Cohen, 1983).

Results

The distribution of attitudes toward academic group work is shown in Figure 1. Fifty one percent of students who completed the survey thought that out-of-class group work should be decreased, while only twenty eight percent thought it should be increased.

Three demographic variables were found to be significantly related to cadet attitudes toward group work: GPA, gender and class. Athletic status and type of major were not significantly related to attitudes. The most significant predictor of attitude was grade point average, with a standardized regression coefficient of -.412. The mean GPA was 2.91 (sd =.56). The inverse relationship between grade point average and attitude towards groups is not irrational; top performers often correctly assume they have the least to gain and the most to lose from working with others. The higher students' grade point averages, the less fond they were of out-of-class group work. The correlation between GPA and attitude toward groups was -.396 ($p < .01$). The next variable found to be significant was gender with a standardized regression coefficient of .259; this suggested female cadets disliked group work more than their male classmates. The mean score for males (M=2.80, SD=1.15), was significantly more positive toward group work than the female participants' average (M=2.05, SD=1.00)($t(104)=-2.70$, $p < .01$). Seniors disliked group work more than juniors. The regression coefficient between class and attitude was -.256 which was significant at the .01 level. Seniors' mean attitudes toward

groups (M=2.44, SD=1.09) was significantly more negative than were juniors' expressed attitudes (M=2.86, SD=1.19)(t(104)=2.31, p<.05). For the overall regression, the R was .535, and the R^2 value was .286, (F(3,101)=14.0, p<.01).

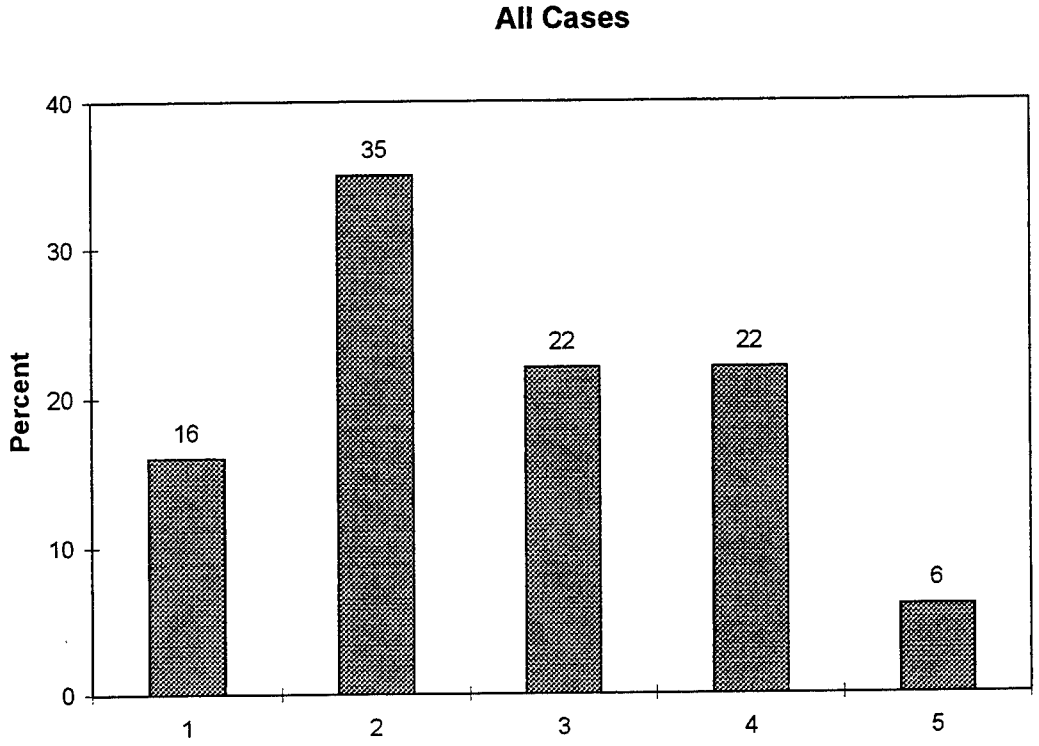


Figure 1
#2. How could the Academy increase educational effectiveness?

1) Greatly decrease the amount of out-of-class group work.
2) Slightly decrease the amount of out-of-class group work
3) Keeping the amount of out-of-class group work the same
4) Slightly increase the use of out-of-class group work
5) Greatly increase the use of out-of-class group work

Discussion

At the Academy, the overall attitude toward group work out of the classroom appears to be slightly negative, meaning that most cadets surveyed feel the Academy could improve educational effectiveness by decreasing the use of out-of-class group work. One possibility is that cadets simply have negative attitudes toward the subjects being taught in which groups are used. This seems unlikely because student groups are used in majors courses as well as core courses, and most students have positive attitudes toward their majors.

Another possible explanation for this apparent dislike of group work is that cadets are resentful of having to work cooperatively in an environment that is inherently competetive. Many cadets see themselves as competitors; they have to be just to gain admittance. Once enrolled, the competition doesn't stop, it only becomes more intense. Cadets are no longer competing with average students, they're competing with other top students from all across the nation. Working together with other cadets means helping them achieve higher grades, and whether it be intentional or unconscious, cadets may resent having to help someone who could be competing for the same pilot slot, medical school or graduate school opportunity. That cadets would be resentful of having to help each other is a somewhat disquieting prospect, but nonetheless, one that cannot be ignored. This situation is not necessarily the fault of the cadets themselves. The Academy environment breeds competitiveness. Rewards go to the people who perform the best and who can make the right impressions on key members of faculty and staff. There seem to be quotas for everything, and limited numbers of the rewards that cadets desire.

There is also the possibility that cadets simply don't respond well to the ways groups are used in the Academy curriculum. The devil is often in the details and the particular constraints of each group project may work against the desired learning objectives. These possibilities are not mutually exclusive. Though classes and assignment procedures themselves seem most culpable, it is still possible that cadets would have difficulty adapting to any cooperative work scenarios within the general competitiveness of the environment. Whatever the reason, the overall negative attitudes toward group work at the Academy are unlikely to be solely attributable to the cadets themselves.

The other findings of this study also pose interesting questions. Students with higher GPAs had more negative attitudes toward group work than students with lower GPAs. Part of the reason may be that students with higher GPAs normally work alone as a way to earn better grades. This also seems to tie in with earlier discussions about competi-tiveness. Perhaps intepersonal competitiveness at the Academy produces better results than cooperation. If so, this would distinguish the Academy from the other institutions where previous research showed academic advantages of cooperation. The overall system may contain incongruities which have effects exactly opposite to their intentions (viz., negative group experiences discourage cadets from developing positive attitudes toward group work).

Other aspects of the results also need to be considered; these include the finding that female cadets were more opposed to group work than male cadets. This result also is contrary to other research which suggests that women and minorities are often the greatest beneficiaries of collaborative pedagogies. One possible explanation is that women at the Academy don't like to work in groups because they are a minority, and likely to be a minority in any groups they work with. Their opinion that education would improve if groups were used less may suggest that women, as a minority, may suffer the most when groups are used inappropriately or ineffectively.

Insight into the finding that seniors disliked group work significantly more than juniors was gained by examining another question on the survey. The survey asked each subject to identify the course in which their most negative and most positive experiences with group work had occured. Responses to the positive experience question were fairly widely distributed across

many different courses and academic divisions. On the other hand, responses to the negative experience question yielded quite an interesting result. A large majority of senior cadets identified a single core Engineering course as their single worst experience with academic group work. This particular class is seen as being almost entirely group-oriented. Most of cadets' time in this class is devoted to working in groups on huge projects that account for less than 50% of their grades. Since this is a junior class, it is usually taken during the junior year (none of the juniors participating in this study had yet completed this course; and only about half were currently enrolled in it). The implication here is that a bad experience in a singe course might explain the significantly increased negativity in seniors' attitudes toward group work.

A large number of cadets participated in this study and the results it yielded were not only significant but have substantial implications. These data show no evidence that the Academy experience instills in cadets a generally positive attitude toward working with others. In fact these data suggest just the opposite: although there are considerable individual differences, the average cadet attitude toward group work is somewhat negative. Particularly distressing are the findings the attitudes amongst the "best" students are the most negative; attitudes grow significantly more negative from the junior to senior year and the women (and perhaps other minorities) appear to be harmed rather than helped by the current misapplication of such an apparantly progressive pedagogy. Perhaps these findings suggest it is critical to not only consider the question of "to group or not to group?" but to pay much more careful attention to the question of "how" and "with what results."

Authors' Note: Views expressed in this paper are those of the authors alone and do not necessarily represent the position or policy of the United States Air Force Academy or any other government agency.

References

Bruffee, K.A. (1987, Mar/Apr). "The art of collaborative learning." Change. pp. 42-47/

Cohen, J. & Cohen, P. (1983). Applied multiple regression/ correlation analysis for the behavioral sciences. Second Ed., Hillsdale, New Jersey: Lawrence Erlbaum Assoc.

Kohn, A. (1986, Sept). "How to succeeed without even vying". Psychology Today. pp. 22-28.

Lyman, Lawrence and Foyle, Harvey C. (1990). Cooperative grouping for interactive learning: students, teachers, and administrators. National Education Association,Washington, D.C.

Porter, D.B., (1989). "Educating from a group perspective: what, why, and how", Proceedings of the Human Factors Society 33rd Annual Meeting. Santa Monica, CA: Human Factors Society. pp. 507-512.

Savoie, Joan M. and Hughes, Andrew S. (November, 1994). "Problem-based learning as classroom solution",Educational Leadership.

Slavin, Robert E. (1987). Cooperative learning: Student teams, 2nd Edition. National Education Association, Washington, D.C.: 1987.

Stahl, Robert J., (March, 1994) "The essential elements of cooperative learning in the classroom", ERIC Digest. Office of Educational Research and Improvement, Washington, D.C.

Indicators of Reflective Thinking In College Faculty

Kate Preston & David B. Porter, DPhil
United States Air Force Academy

Fifteen Air Force Academy faculty members completed a reflective judgment assessment questionnaire. Demographic and instructor critique information were also collected. Current literature suggests age, academic discipline and level of education are strong predictors of reflective judgment level. This study sought to extend our understanding of reflective judgment by assuming classroom teaching was a real-life, ill-defined problem and that student ratings would be related to higher levels of reflective judgment. Although the study did find small differences in expected directions for academic discipline and degree, no relationship between reflective judgment level and student ratings was found.

William Perry's book, Forms of Intellectual and Development in the College Years, (1970) presents a cognitive foundation for adult intellectual development. His theory describes the development of thinking through the distinct stages of Dualism, Early and Late Multiplicity, and Contextual Relativism. Patricia King (1992) outlines a similar progression of cognitive development in the Reflective Judgment Model. This model is presented as a heuristic that, like Perry's, suggests a progression of intellectual development that affects the way information from the world is perceived and employed to solve problems and make decisions.

Reflective judgment development is segmented into seven stages. In Stages 1 & 2 (pre-reflective), knowledge is gained through direct, personal observation or the word of authority. The view of knowledge in Stage 3 assumes answers are absolute but, unlike previous stages, allows that some answers may be temporarily inaccessible. People in this stage experience a great deal of difficulty with ill-defined problems. Stage 4 reasoning shows an appreciation of the importance of evidence; individuals in this stage recognize the difference between well and ill-defined problems but may have difficulty dealing with ambiguity. With Stage 5 reasoning, individuals view specific dilemmas in broader contexts and identify specific factors which cause difficulty but then have difficulty resolving these issues and drawing conclusions. Stages 6 & 7 represent the highest level of reflective judgment and are characterized by an understanding the importance of knowledge context, as well as its active construction and interpretation (King, 1992).

The relative ability to frame and resolve ill-defined problems is the definitive characteristic in assessing levels of reflective judgment. As such, metrics associated with the model often present an ambiguous problem, provide an outline for a general solution, and assess a subject's response against various reflective thinking criteria (Lynch, 1995). Among these criteria are acknowledgment of the "ill-defined" nature of the problem, recognition of personal bias and its influence, use of a guiding principles, modification of solutions when inconsistencies arise, and awareness of the relationship between problems and their context. Evidence of these processes are indicative of higher levels of reflective thinking, whereas, the failure to incorporate these activities reflects less development.

Each semester college faculty are presented with a particularly ambiguous dilemma. There is no approved solution for teaching a college course. Both teachers and students have preferences for different teaching styles and teachers must accommodate an array of thinking and learning styles, systematically varying the approach relative to each student and topic (Sternberg, 1994). Educators who are themselves at higher levels of reflective thinking should be better equipped to accommodate student variety and, through awareness of context and acceptance of ambiguity, be able to encourage reflective thinking in all of their students (Strange, 1992). Intuitively, it seems this ability would be present in all teachers. However, there has been limited research in this area.

To adequately understand the implications of reflective thinking in college educators, the possible indicators of development level need to be examined. Pilot research done by Wood, et. al. (1994) indicates that education level and reflective thinking are strongly correlated. The current study focuses on five possible correlates of reflective thought: age, level of education, academic discipline, gender, and perceived instructor effectiveness. The first four of these we assumed to be independent predictors of a faculty members' level of reflective judgment while the fifth, perceived classroom effectiveness, is assumed to be more a consequence than a cause of intellectual complexity. Recognizing which factors are related to reflective judgment could have important implications for the selection of teachers, design of curriculum, and educator training.

Method

Participants

Fifteen faculty from the United States Air Force Academy who had recently participated in a study on correlates of perceived teaching effectiveness (Porter & Benson, 1995) volunteered to participate in this study. Twelve were Air Force officers, the remaining three were civilian professors. There were two female and thirteen male instructors whose average age was 39 years. Sixty percent of the sample held doctorate degrees; the others had earned Masters degrees. Professors from seven distinct academic disciplines were sampled; the number of faculty from each department were: 4 behavioral science, 1 military arts and science, 2 computer science, 5 biology, 1 legal studies, 1 foreign language, and 1 English. Questionnaires from two other subjects were not returned.

Apparatus

The instrument used in the current study was adapted from the Reflective Judgment Appraisal (RJA, 1992, Version P/Revised: 1993). This questionnaire is designed to assess thinking about ill-defined problems. Levels of performance are based on the Reflective Judgment Developmental Model (King, 1992). Subjects were presented with a nature versus nurture controversy and asked to resolve this issue within an ill-defined context. The appraisal was not dependent on the content of subjects' answers; however, four basic characteristics of their thinking were considered: 1) opinion justification, 2) degree of certainty, 3) appropriate criteria for judgment, and 4) explanation for expert disagreement. In each of the four sub-sections, subjects are asked to use a four-point scale to rate the degree to which alternative statements matched their own thinking. Subjects were then asked to rank order the top three statements from each group which most

closely matched their own views (Wood, 1994). Questionnaires were scored by assigning stage utilization scores based on their rank ordering of the statements. Composite scores were the arithmetic average of the four topic scores.

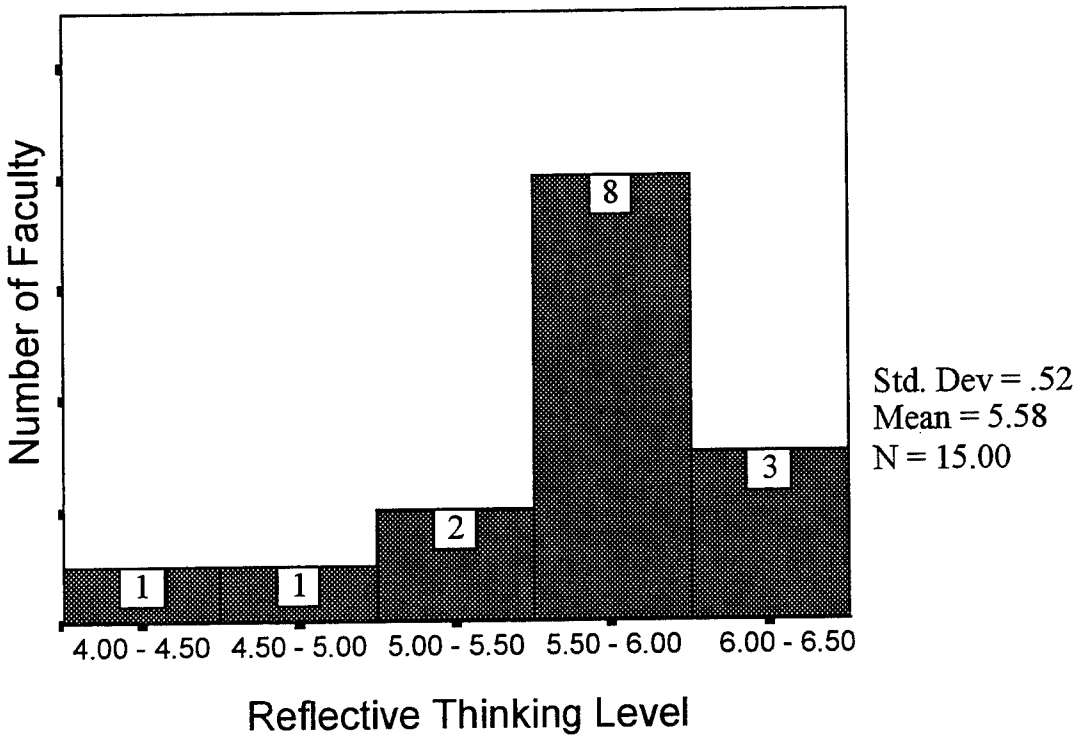
Procedure

Seventeen faculty members responded to an invitation to participate in this study. Each participant was asked to complete the eight page questionnaire, which included the adapted Reflective Judgment Appraisal and demographic information. Two questionnaires were not returned; the remaining fifteen were received through the Academy distribution system within a few days. Instructor effectiveness data from the previous study (Porter & Benson, 1995) were used as an additional independent variable. Individual feedback, as well as copies of this report, were made available to participants.

Results

The instructor effectiveness critique scores were the sum of 5 items from the Academy's annual instructor critique. Scores ranged from 4.32 to 5.58 on a six-point scale with a mean score of 5.12 (SD = .36). Eight of the participants instructed technical courses (e.g., computer science and biology), while 7 were affiliated with non-technical disciplines (e.g., behavioral science, English, etc.). The reflective thinking levels of Air Force Academy instructors ranged from 4.25 to 6.22. The mean level of reflective thinking was 5.58, with a standard deviation of .52

Table 1: Reflective Thinking Level in USAFA Academic Faculty (% of sample)



Pearson correlations between RT Levels and demographic variables showed a limited relationship. Age and education level, as well as academic department and critique rating, were significantly related. However, none of these correlations with RTL were found to be significant. The strongest relationships were found between and RT level and level of education (r=.42, p=.12) and RT level and academic department (r=.38, p=.20).

Table 2: Correlation table of RT Level, Age, Academic Department, Level of Education, and Instructor Effectiveness Rating

Factors:	Age	AcDept	EdLevel	Rating
RT Level	.049	.377	.419	.044
Age		.063	.468*	.256
AcDept			.327	.543**
EdLevel				.251

*p<.1 **p<.05

Age, education level, and academic department were each grouped with reflective thinking level to create new independent variables (Age*RT, Ed*RT, and Dept*RT). Multiple regression analysis was used to examine potential interactive effects; none were significant. Two combined factors, reflective thinking level factored with education level and academic department, showed the strongest relationship. Although the strength of the relationships were increased from those found by the simple correlations, no significance was found.

A direct examination of the effects of two independent variables (academic division and education level) on RT Level was conducted using independent t-tests. Although the mean RT Level for the 8 non-technical faculty members (m=5.76, sd=.26) was higher than the mean RT Level for the 7 technical faculty members (m=5.38, sd=.68), the difference was not significant (t=1.39, p=.20). However, Levene's Test for Equality of Variance (F=5.06, p<.05) showed there was significantly more variance in the technical faculty members than among the non-technical faculty. A similar result was found when comparing the 9 faculty members with doctorate degrees to the 6 with masters degrees. Although the mean RT Level for faculty with doctorates (m=5.75, sd=.24) was higher than for those with masters degrees (m=5.33, sd=.72), the difference was not significant (t=1.66, p=.12). Once again, Levene's Test (F=8.42, p=.01) suggested the differences in variance in the two groups was significant: the six individuals with masters degrees showed much greater variability.

Discussion

The current study failed to support existing research on reflective thinking. Small but insignificant differences were found in average reflective judgment scores based on academic discipline and degree. Fifteen subjects do not provide the statistical power to find anything but

very large effects. Perhaps more interestingly, however, significant differences were found in the amount of variance in some groups. The virtual absence of variability in the scores of those with doctoral degrees or those teaching in non technical disciplines, might suggest ceiling effects. Although the instrument employed in this study had been successfully piloted with students and larger samples of faculty, these results suggest that the instrument itself may have insufficient high end discriminability. In particular, the sample used in this study had first volunteered to participate in a study designed to examine the predictors of perceived classroom effectiveness (Porter & Benson, 1995). Having already received individualized feedback from this study, they then agreed to participate in this further study of the relationship between reflective judgment and classroom effectiveness. Together these choices reflect an attitude toward inquiry characteristic of the highest Reflective Judgment Levels. Thus the fact that so many of them scored at this level should not be a surprise.

In the broader context, the failure of this study to identify significant demographic predictors of RT Level and the apparent lack of relationship between RT Level and student perceptions of teaching effectiveness do not necessarily suggest that no relationships exist. The fact that 13 of 15 faculty members received reflective judgment scores of 5.0 or higher might be interpreted to mean that all of them had sufficient insight to understand the process of classroom teaching but may have lacked other skills or attributes needed to be perceived as effective teachers. Personal warmth, clarity of expression, sensitivity to students, timeliness, organizational ability are among the many attributes associated with high student ratings that may not be tied to reflective judgment. As is often the case, more research with the specific instrument used in this study as well as the role of reflective judgment in teaching and learning is needed. Our hope is that this study makes a small contribution to this larger quest.

References

Colton, A, B. & G. M. Sparks-Langer. (1993) A Conceptual Framework to Guide the Development of Teacher Reflection and Decision Making. Journal of Teacher Education, 44, 45-54.

King, P. M. (1992) How do we know? Why do we believe?: Learning to make reflective judgments. Liberal Education, 78, 2-9.

Lynch, C. L. (1995) Assessing reflective judgment (unpublished manuscript).

Perry, W. (1970). Forms of intellectual and ethical development in the college years. New York: Holt Rinehart and Winston.

Porter, D, B. & M. Benson. (1995) Correlates of course and faculty perceived effectiveness. In (S. M. Eisenhut, Ed.), United States Air force Academy Educational Outcomes Assessment Working Group; Phase I: Final Report. US Air Force Academy, CO: Dean of the Faculty.

Sternberg, R, J. (1994, Nov) Allowing for thinking styles. Educational Leadership.

Strange, C. (1992) Beyond the classroom: encouraging reflective thinking. Liberal Education, 78, 28-32.

Wood, P.K., King, P., Kitchener, K.S., Lynch, C. (1994). Technical manual to accompany the reflective thinking appraisal (RTA; Ver. 1.0). Reflective Thinking Associates, 1-25.

Focus Group Technique as a Classroom Learning Activity

Lt Col Bernard Asiu[1]
Armstrong Laboratory, Human Resources Directorate

Abstract

This study reports on a successful application of the focus group technique to support student learning outcomes for a course in the psychology of adult learning. Two focus groups with six students each were conducted around student experiences using a case-based lesson planning software tool. Analysis of focus group audiotapes indicate students were able to participate at the application, analysis, synthesis and evaluation levels of Bloom's taxonomy. Focus group and classroom discussion techniques are compared and discussed. There are several distinctions to focus groups which the literature supports as positively related to student participation and learning. These include the structured nature of planning, conducting and analyzing focus groups, participant diversity, the nonevaluative component to focus groups, and the focus on participant interaction to develop a shared perspective. For educators interested in experimenting with pedagogy, this study provides considerations to improve their practice and understanding of teaching.

Introduction

The focus group technique has often been used in education and educational research to answer questions regarding needs assessment, program development, implementation and evaluation. A recent review of ERIC (1992 - 1995) and PsychLIT (1990 - 1995) databases reveals over thirty activities involving the use of focus groups to address such topics as education and curriculum reform, classroom practice, teaching excellence, college entry experience, strategic planning, and program development for special needs (e.g., adolescent suicide, AIDS, handicapped). Participants in these focus groups include major stakeholders such as students, parents, faculty and supporting staff members. There are doubtless many applications of the focus group technique to education that do not find their way into the research literature. Despite the success of the focus group technique in identifying participant perspectives to education-related interventions or concepts, the technique has not yet found its way into the classroom as a teaching strategy. However, Bogdan and Biklen (1992) discuss pedagogical uses of qualitative research to improve educator effectiveness, to enrich teacher training and to add participant observation and communication skills to school curriculums. This study reports on a successful effort to extend the focus group technique from traditional evaluation and research applications to support pedagogy for a graduate course in the psychology of adult learning.

[1]Special acknowledgement to Dr. Robert Tenneyson for the opportunity to conduct the focus groups and to Dennis Gettman and Chuck Swanberg as focus group moderators. The views expressed herein are solely those of the author and do not reflect official views of the the United States Air Force or the Armstrong Laboratory.

Method

The focus group activity was conducted at the University of Minnesota in a graduate-level course on the psychology of adult learning. The activity was designed to support course objectives related to adult learning theories, instructional design, curriculum development, technology application, and training evaluation. The focus group exercise centered around student experiences using GUIDE (Guide to Understanding Instructional Design Expertise). GUIDE is a case-based lesson planning software tool to help novice training developers organize lesson content to better support learning. GUIDE is based on the theoretical work of Gagné and is organized around his model for the Nine Events of Instruction (e.g., gain attention, describe the goal, provide learning guidance, assess performance, enhance retention and transfer) (Gagné, Briggs & Wager, 1992).

Students were given a review of GUIDE developmental history and theoretical approach with particular emphasis on the use of Gagné's Nine Events of Instruction and case-based reasoning to support lesson design. Students were next shown a brief demonstration of GUIDE and allowed forty minutes for unguided exploration of the software. After completing the laboratory activity, students were assigned to two equally represented interview groups based on previously obtained self-reports of knowledge of instructional systems development, teaching and training experience, and experience in developing computer-based training materials. Students were reminded that discussions would be audiotaped and were briefed on the structure and format of the focus group interview. The focus group moderators facilitated discussion around six topic questions regarding student response to the GUIDE approach to teaching and learning. After the focus group activity, students returned to the classroom for a review of focus groups as a data gathering technique and to share their focus group experiences through discussion.

Results

The focus group audiotapes were transcribed and individual statements grouped into coherent general observations or themes. Only a single researcher participated in the data reduction and interpretation. Overall, students were very engaged and excited in sharing their GUIDE experience during the focus group interviews. Student-to-student interaction was particularly active and high quality. In general, students understood and argued the complexities of technology applications to learning and pedagogy better than they gave themselves credit for. There were several instances where a student would put forth a critical comment to the moderator only to have other class members answer in animated discussion. In one instance, a student wanted to know what value others saw in computer-based training. The ensuing discussion between students uncovered over a dozen benefits to computer-based training that one might only find in a good authoritative text on the topic. In another instance, a student asked why access the Nine Events via computer instead of using traditional paper media. Responses from other students pointed out the value of dynamic linking and the ability to efficiently edit, catalog and search material. These same students then began to discuss how using GUIDE changes the whole nature of lesson development, pointing out the shift in activities (planning, organizing, building, backtracking, testing, revising) that would likely occur from using GUIDE.

The focus group activity was also valuable in helping students to meet course objectives. For example, with regard to the conceptualization of computer technology applications for the design and delivery of material in adult learning, students learned that previous experience with computers and computer-based training is related to the way people perceive the technology:

> Novice computer users felt overwhelmed by the interface (media, buttons and hypertext). They felt "lost" in not always knowing "where they've been and where they're going." One student suggested a "where am I button." Another student with extensive exposure to computers felt GUIDE was "too basic, I quickly became bored with it." This same student provided ideas to make GUIDE more flexible including the capability to keep the notepad available as a smaller window in the active screen to keep from "bouncing to another screen just to make notes on something you're looking at in the current screen."

With regard to goals for the elaboration of Gagnés' Nine Events of Instruction and case-based reasoning in an adult learning context, students showed strong integration of the Nine Events of Instruction into existing schemas for successful learning:

> "The good instructors I remember did all these things (Nine Events)." Students commented that the Nine Events match their own personal experience on what works, it seems like "common sense." The simplicity, structure and checklist nature of the Nine Events can help new teachers feel less overwhelmed in their new environment.

Students showed the capacity to evaluate the application of case-based reasoning to lesson development:

> "GUIDE (pedagogy) focuses on the presentation of instruction (between computer and student). How can it be used when you want students to develop skills on their own like in exploratory learning or small group activities or discussions?" Another student disagrees, "You can design your lesson as a discussion. The goal for performance would be a discussion but you still need to establish the goal and to motivate."

Discussion

The focus group outcomes indicate students were able to participate at the application, analysis, synthesis and evaluation levels of Bloom's taxonomy. To the extent that this higher level cognitive learning can be attributed primarily to the focus group activity is difficult to determine without comparison groups using traditional classroom discussion techniques. However, it might be helpful to compare the focus group and classroom discussion to see how any similarities and differences might be related to the student outcomes in this study. The focus group and discussion techniques are compared in Table 1 using criteria adapted from Wilen (1990) to discuss various forms of classroom discussion.

	Interaction pattern	Level of thinking	Types of interactions	Degree of structure	Leadership style	Goal
Discussion	1) teacher - student 2) student - student	lower & higher level cognitive learning	questions, statements, acknowledg e-ments, silence	low	teacher as director	improve understanding and analysis
Focus Group	1) student - student 2) teacher - student	higher level cognitive learning	questions, statements, acknowledg e-ments, silence	high	teacher as facilitator	shared understanding from participant interaction

Table 1. Comparison of focus groups and classroom discussion technique on Wilen's (1990) dimensions of classroom discussion.

Although the stages of the activity and the roles of teachers and students are similar in both techniques, the nature of the interaction is distinctly different. Student to student communication is emphasized in focus groups due to the stated goal to develop a common understanding from shared perspectives. Student initiated responses in classroom interaction typically accounts for less than 10% of student talk (Klinzing & Eurich, 1988). Wong (1991) suggests that teachers' power role in the classroom may decrease student participation due to perceptions on the student's part for a "correct answer" and from students' reluctance to be evaluated in the presence of their peers. Although the relationship between student to student interaction and discussion quality is unknown, an hypothesis might be that the lack of an evaluative component to focus groups encourages participation and supports learning outcomes. It is interesting to note that the type of interactions (questions, non questions, wait time etc.) found in both techniques is generally the same and have been shown effective in increasing the length and quality of student responses (Dillon, 1990).

Another key difference relates to the degree of structure in planning and conducting the two activities. Klinzing & Floden (1990) describes that the openness of the classroom discussion technique decreases teachers' reliance on planning. On the other hand, the focus group technique imposes a deliberate structure which may help address reasons why teachers do not use discussions more often. These include not directly addressing critical thinking skills as course outcomes, unwillingness to invest the time, and reluctance to relinquish control (Kindsvatter, 1990). That the students in this study were deliberately assigned to create two heterogeneous focus groups in terms of prior teaching / training experience and familiarity with computer-based training may also be related to the high levels of learning observed. Gall & Gall (1990) reviews that heterogeneous membership facilitates student interaction in discussion groups to develop student's capability for moral reasoning. Membership diversity along dimensions deemed important to the focus group topic is important to foster participant interaction. Therefore, it might be hypothesized that the diversity structured into this particular focus group is also related to improved learning. A final aspect to consider is that the structured GUIDE review and

exploration prior to the focus group meeting may have provided the background knowledge and shared experience to encourage participation and learning. Wong (1991) shows that discussions are enhanced when students can personally relate to and share in the topic of discussion.

Conclusion

Although the absence of a classroom discussion comparison group makes it difficult to discern the extent to which the focus group technique contributed to the positive learning outcomes observed in this study, there are several distinctions to focus groups which the literature supports as positively related to student participation and learning. These include the structured nature of planning, conducting and analyzing focus groups; participant diversity, the nonevaluative component to focus groups; and the focus on participant interaction to develop a shared perspective. For educators interested in experimenting with pedagogy, this study provides considerations to improve their practice and understanding of teaching.

References

Bogdan, R., & Biklen, S. (1992). Foundations of qualitative research for education: An introduction to theory and methods (2nd Ed.). Boston: Allyn Bacon.

Dillon, J. T. (1990). Conducting discussions by alternatives to questioning. In William W. Wilen (Ed.), Teaching and learning through discussion, (pp. 79-96), Springfield, Il: Charles C. Thomas Publisher.

Gagné, R. M., Briggs, L. J., & Wager, W. W. (1992). Principles of instructional design (4th Ed.). New York: Harcourt Brace Jovanovich.

Gall, Joyce P., & Gall, Meredith D. (1990). Outcomes of the discussion method. In William W. Wilen (Ed.), Teaching and learning through discussion, (pp. 25-44), Springfield, Il: Charles C. Thomas Publisher.

Kindsvatter, Richard (1990). Teacher social power and classroom discussion. In William W. Wilen (Ed.), Teaching and learning through discussion, (pp. 113-126), Springfield, Il: Charles C. Thomas Publisher.

Klinzing, Hans G., & Floden, Robert, E. (1990). Learning to moderate discussions. In William W. Wilen (Ed.), Teaching and learning through discussion, (pp. 175-202), Springfield, Il: Charles C. Thomas Publisher.

Klinzing, Hans, G. & Klinzing-Eurich, Gislea (1988). Questions, responses and reactions. In J. T. Dillon (Ed.), Questioning and discussion: A multidisciplinary study, (pp. 212-239). Norwood: Ablex Publishing.

Wilen, William W. (1990). Forms and phases of discussion. In William W. Wilen (Ed.), Teaching and learning through discussion, (pp. 3-24), Springfield, Il: Charles C. Thomas Publisher.

Wong, E. David (1991). Beyond the question / nonquestion alternative in classroom discussion. Journal of Educational Psychology, 83(1), 159-162.

Journal Writing: Its Effects on Objective Test Performance in an Upper Division Leadership Class

Robert C. Berger
Gary A. Packard, Jr.
Craig A. Croxton
USAF Academy

Abstract

Numerous studies have shown no effect of journal writing on objective test scores. Most of these studies have used overall test scores as measures of journal-writing effectiveness. This paper reports the results of a study that examined performance on specific test questions and related journal entries. We found that students who wrote journal entries on topics related to specific test questions were more likely to correctly answer those objective test questions on the final exam than students who did not write on the topic.

Instructors, teachers, professors, and educators at all levels struggle to make formal education a precursor to life-long learning. At the United States Air Force Academy (USAFA), one of the educational outcomes states, "We want to develop an attitude of intellectual curiosity in our graduates that predisposes them to lifelong learning" (italics in the original).

Many instructors have attempted to encourage critical thought, application of knowledge, and an attitude of intellectual curiosity through journal writing. The pedagogical benefit of journal writing is touted by many. The Journal Book (Fulwiler, 1987) contains many examples of the ways teachers use journals to improve student learning in such diverse disciplines as physics, chemistry, political science, and geography. Fulwiler states, "Journals are useful tools for both students and teachers. They can help students prepare for class discussion, study for examinations, understand reading assignments, and write formal papers" (p. 6).

The uses of journals in education are as varied as the instructors who use them. However, support for the educational value of these writing assignments seems almost universal. Yinger (1985) discusses many different types of journal exercises and concludes "writing is a powerful tool for learning as well as for communicating" (p. 31). Brodsky and Meagher (1987) report using journals for up to 75% of the course grade in a political science course. The journals had specific requirements and were collected frequently. They found that expressive or exploratory writing dominated the journals, and concluded that the journals improved student writing and learning by providing students different avenues to apply lessons, ask questions, and improve analytical capabilities. Grumbacher (1987) used journals in an introductory physics class and found students who were able to connect ideas in their journals (i.e., how concepts apply to their own world) were better problem solvers. Hettich (1990) also supports the use of journals as an effective avenue to enable students to relate course ideas to their experiences.

However, most of the support for journals comes from anecdotal reports and not through empirical data. McCulley (1986) has stated that the literature is replete with general speculations about the effectiveness of writing on learning, but we know little about the specifics of writing to learn. For example, we do not know what content can be taught through writing (McCulley, 1986). In fact, the impact of journal writing on learning as measured by the results of classroom testing has yielded less than encouraging results. Day (1994) divided an introductory sociology class into two random groups. One group received points based on attendance while the other group received points for keeping a journal on course material. Her study found whether a student completed the journals or merely was required to attend class did not predict essay or multiple-choice scores. Harchelroad and Rheinheimer (1993) used journals in a summer math course and found the students in the control group did better than the students who wrote journals. An interesting exception to this was the students with the lowest entering math skills who wrote journals scored as well as the non-journal writers. Jensen (1987) used journals in a year-long physics course. For the fall semester he required students in one section to keep a journal whereas students in the other section did not. For the spring semester, he reversed the requirements for the two sections. He found no differences in test scores at the end of either semester on objective, problem solving examinations. Selfe, Petersen, and Nahrgang (1986) reported the results of an experiment using journals in a ten week long college-level math course. One of the sections used journal and tests, one used quizzes, and one section used only tests as evaluations. All three sections took the same tests. Selfe et al. found no differences in objective test scores among the three sections.

Assuming we can accurately measure student learning through classroom testing, the above results seem to indicate journal writing is not improving student learning as measured by objective test scores. As Hettich (1980) notes, reading and commenting on student journals is a time consuming task for the instructor. If student learning, as measured by classroom testing, is not improved through journal writing, what is being gained by the time and effort being spent by both teacher and student?

One consistent approach to the research on the effectiveness of journal writing on objective test measures is that journal writing has been compared with overall test scores, not on correct answers for individual test items. Britton, Burgess, Martin, McLeod, and Rosen (1975) suggest when people write about new information they learn and understand the information better. If this is true, then perhaps the learning needs to be analyzed on a more basic level, such as looking at individual test questions. We wondered if the results might be different if we compared a journal entry written on a specific topic with a test item on the same topic. For example, if a physics student wrote a journal entry on friction, would the student be more likely to answer test questions related to friction correctly than a student who had not written on friction? Our study was designed to look at this level of analysis using a leadership course currently taught at USAFA. We hypothesized that writing on a given topic would lead to higher scores on examination questions on the same topic.

Method

The Behavioral Sciences and Leadership Department at USAFA teaches a semester-long (seventeen week) junior/senior level course entitled "Leadership Concepts and Applications." This course is currently taken by approximately two thirds of cadets at USAFA. One of the course requirements is the writing of leadership application papers. These papers are reflective journals in which the students are required to produce 6-10 entries, each approximately three pages in length. The purpose of the entries is to encourage students to relate course concepts to their experiences and reflect on the lessons learned from these experiences. For example, on the topic of communication, the student might write about her commander's poor use of communication in the squadron, analyze what was wrong with the communication using a systems model of communication discussed in class, and then discuss how she would improve on the communication if she were the commander. The entries were graded pass/fail. A passing grade was given to an entry which accurately reflects course content, shows depth in reflection, and was relatively free of grammatical and spelling errors. All students were allowed one attempt to re-write a failing entry to a passing level. Students chose from 16 topics and were required to write 10 passing entries to receive 100% on their journal grade for the semester, pass nine entries for 95%, pass eight entries for 90%, and so on, down to five passing entries for 65%. Five passing entries was the minimum number needed to pass the course. The overall journal grade accounted for 20% of the grade in the course.

Participants

Participants were 234 students enrolled in the leadership course during the spring semester 1995.

Design and Procedure

Instructors were asked to keep a log of their students' entries during the semester. Each of these entries was written on a specific topic, such as communications or conflict. Some of the papers were not on a testable course topic (i.e., a paper on a leadership issue in the news) and these papers were not used in the study. We then classified test questions from two course examinations and the final examination based on the topic addressed by the question. If a question covered more than one topic or was related to a topic not included among the 16 journal topics, then that question was not used in the analysis. All test questions were multiple choice or true/false questions. Eight topics had both a paper and at least one related test question. The topics were 1) The Leader-Follower-Situation Model, 2) Power, 3) Communication, 4) Conflict, 5) Values, 6) Motivation, 7) Stress, and 8) Contingency Theories of Leadership.

Results

For each test question, students could be categorized into one of four groups: 1) wrote on the topic and correctly answered the question; 2) wrote on the topic and incorrectly answered the question; 3) did not write on the topic and correctly answered the question; 4) did not write a paper on the topic and incorrectly answered the question. We evaluated the resulting 2 x 2 contingency table using the Chi-square test to evaluate the relationship between journal writing and performance on related test questions.

Exams 1 and 2 showed that there was no relationship between whether the student had written on a journal topic and the probability of correctly answering a related test question (Exam 1: $\chi^2 = 2.049$, $p > .05$, Exam 2: $\chi^2 = 0.056$, $p > .05$). The final exam, however, did show that there was a significant relationship between writing a journal entry and getting a related test question correct (Final Exam: $\chi^2 = 17.999$, $p < 001$). We then partitioned the final exam data into two parts: 1) questions relating to journals written early in the course (labelled Final $\text{Exam}_{\text{Journals 1-5}}$ in Table 1), and 2) questions relating to journals written later in the course (Final $\text{Exam}_{\text{Journals 6-8}}$). Chi-square tests showed that the relationship between writing a journal entry and correctly answering a related test question was significant only in the early journals ($\chi^2 = 6.942$, $p < .01$) and not in the later journals ($\chi^2 = 2.312$, $p > .05$). See Table 1 for a summary of the data analysis. Table 2 lists percentages of questions answered correctly or incorrectly classified by whether the student had written the journal or not.

Table 1. Chi Square Summary by Evaluation Type

	df	N	χ^2	p-value
Exam 1	1	2343	2.049	> .05
Exam 2	1	2167	0.056	> .05
Final Exam	1	5251	17.999	< .001
Final Exam Journals 1-5	1	2975	6.942	< .01
Final Exam Journals 6-8	1	2276	2.312	> .05

Table 2. Percentages of Questions Answered Correctly or Incorrectly by Exam

	Journal & Correct Answer	Not Journal & Correct Answer	Journal & Incorrect Answer	Not Journal & Incorrect Answe
Exam 1	58.8%	55.5%	41.2%	44.5%
Exam 2	80.6%	80.1%	19.4%	19.9%
Final Exam	68.1%	62.4%	31.9%	37.6%
Final Exam Journals 1-5	71.5%	66.7%	28.5%	33.3%
Final Exam Journals 6-8	62.1%	59.0%	37.9%	41.0%

Discussion

For Exams 1 and 2 our data are consistent with previous studies showing journal writing has no effect on objective test performance. However, the data from the final exam show an increase in performance for those students writing journals. One difference between the final exam data and the data from Exams 1 and 2 is the increased time interval between writing the

journals and the objective testing. In an attempt to see if the time interval affected the data we divided the final exam data into two groups, journals written prior to Exam 1 and those written after Exam 1. Our analysis of this partition suggests there may be a relationship between journal writing and objective testing if the time interval between writing and testing is long enough. At the time of the final exam it had been almost ten weeks since students wrote on the first five journal topics. The studies cited in the introduction evaluated journal writing in courses where the interval between journal writing and testing was much shorter.

It is not clear why the extended interval between writing and testing in this study may have contributed to the superior performance on objective test questions of students who wrote on a related topic (71.5%) compared to those students who did not write on the topic (66.7%), but one possible explanation may be the self-referent effect (Rogers, Kuiper, & Kirker, 1977). Rogers et al. found that when subjects evaluated word lists using structural, phonemic, semantic, or self-referent (asking if the word describes themselves) tasks, recall performance for the words in the self-referent conditions were much higher than in any of the other conditions. Rogers et al. suggested that this type of encoding results in more enduring memory because of the initial depth of processing. Students in this course were encouraged to write entries related to their personal experience and this may have resulted in deeper processing for the journal topics than for those students who only read and listened to lectures on the same topics. The deeper processing, and more enduring memories, only became an advantage at the longer time interval between writing and testing, as revealed in the final exam. Future research should further examine this relationship between the time of writing and the time of objective testing. Research is also needed to explore whether the self-referent effect is a factor in the improved recall for objective test questions shown in this study. If the self-referent effect does influence performance,we would expect to find that a different type of journal assignment (one that does not require a connection to the student's personal experience) may not influence performance on objective tests.

References

Britton, J., Burgess, T., Martin, N., McLeod, A., & Rosen, H. (1975). The development of writing abilities. London: MacMillan.

Brodsky, D., & Meagher, E. (1987). Journals and political science. In T. Fulwiler (Ed.), The journal book (pp. 375-396). Portsmith, NH: Boynton/Cook.

Day, S. (1994). Learning in large sociology classes: Journals and attendance. Teaching Sociology, 22, 151-165.

Fulwiler, T. (1987). The journal book. Portsmith, NH: Boynton/Cook.

Grumbacher, J. (1987). How writing helps physics students become better problem solvers. In T. Fulwiler (Ed.), The journal book (pp. 322-329). Portsmith, NH: Boynton/Cook.

Harchelroad, J. L., & Rheinheimer, D. C. (1993). Journal writing: An analysis of its effectiveness in a college-level developmental mathematics class. Research and Training in Developmental Education, 9, 55-63.

Hettich, P. (1980). The journal revisited. Teaching of Psychology, 7, 105-106.

Hettich, P. (1990). Journal writing: Old fare or nouvelle Cuisine? Teaching of Psychology, 17, 36-39.

Jensen, V. (1987). Writing in college physics. In T. Fulwiler (Ed.), The journal book (pp. 330-336). Portsmith, NH: Boynton/Cook.

McCulley, G. A. (1986). Research in writing across the curriculum. In A. Young & T. Fulwiler (Eds.), Writing across the disciplines (pp. 42-48). Upper Montclair, NJ: Boynton/Cook.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. Journal of Personality and Social Psychology, 35, 677-688.

Selfe, C. L., Petersen, B. T., & Nahrgang, C. L. (1986). Journal writing in mathematics. In A. Young & T. Fulwiler (Eds.), Writing across the disciplines (pp. 192-207). Upper Montclair, NJ: Boynton/Cook.

Yinger, R. (1985). Journal writing as a learning tool. Volta Review, 87 (5), 21-33.

No Pain, No Gain:
The Effect of an Intelligent Tutoring System on F-15 Troubleshooting Performance

Bradley S. Boyer, Ellen P. Hall, Anna L. Rowe, and Robert A. Pokorny
Air Force Armstrong Laboratory
Fifteenth Applied Behavioral Sciences Symposium

Abstract

This study compares the effects of two intelligent tutoring systems on the troubleshooting performance of crew chiefs who maintain the hydraulic subsystems of the F-15. Fifty-one F-15 crew chiefs assigned to Tyndall and Elmendorf Air Force Bases were assigned to one of two tutor groups or to a control group. Pre- and post-test scores on a verbal troubleshooting test indicate only one of the tutors significantly improved troubleshooting performance, and that the effect was most pronounced when the troubleshooting scenario required students to develop their own troubleshooting strategies. This paper examines possible sources of this effect, including the instructional features of the tutor which produced it.

The purpose of this study was to evaluate and compare the instructional effectiveness of two intelligent tutoring systems that teach troubleshooting on the hydraulic subsystems of F-15 aircraft. The first tutor, "Hydrive," is a research and development prototype developed under the Armstrong Laboratory's Basic Job Skills (BJS) Program by Educational Testing Service (Gitomer, Steinberg, and Mislevy, in press; Steinberg and Gitomer, 1994). The second tutor, the F-15 Pneudraulics Tutor," was developed by Galaxy Scientific Corporation with funding from the Air Force Human Systems Center Program Office under the Maintenance Skills Tutor (MST) program. As the program responsible for "productizing" the training technologies developed under laboratory programs like BJS, the MST program was interested in evaluating ways of reducing the costs of developing intelligent tutoring systems.

Both Hydrive and the F-15 Pneudraulics Tutor are grounded in an instructional philosophy based on principles of apprenticeship learning: the goal of the tutors is to provide practice on complex problems with the support of an intelligent "coach." Thus, both tutors provide supported, simulation-based troubleshooting practice through a variety of troubleshooting scenarios which are, in fact, identical in the two tutors. The differences between the tutors are reflected in the types of support capabilities they provide and can be most easily described in terms of a model of skilled problem solving described by Gott (1990):

> Whether operating a word processor or diagnosing a faulty engine, the human performer is required to select and execute procedures to interact with an object to achieve a set of goals. The knowledge and processes that constitute that performance are (a) procedural (or how-to-do-it) knowledge; (b) declarative (or domain) knowledge of the object (often called system or device knowledge); and (c) strategic (or how-to-decide-.what-to-do-and-when) knowledge (p. 100).

Whereas Hydrive emphasizes the declarative and strategic knowledge underlying task performance, .the F-15 Pneudraulics Tutor emphasizes the declarative and procedural knowledge components. This paper describes the results of a comparative evaluation of the two tutors and interprets them in terms of the differences between the tutors' instructional features, or more specifically, in terms of the different types of instructional support provided to students in the two tutoring environments.

Method

Participants

Participants were 51 Air Force technicians in the F-15 Crew Chief career field. The technicians all worked directly with hydraulics equipment and were selected to achieve a range of proficiency within the Air Force five-level skill classification system. Forty technicians were assigned to receive tutoring from one of the two tutors. This assignment was achieved by first matching the technicians on the basis of three measures: (a) an assessment of pretest troubleshooting performance, (b) ASVAB mechanical score, and (c) months of troubleshooting experience. Members of each matched pair were then randomly assigned to receive tutoring from either Hydrive or the F-15 Pneudraulics tutor. The remaining eleven technicians participated as control subjects and received no tutoring.

Materials and Procedure

Technicians' troubleshooting knowledge was assessed at pretest and at posttest using two versions of a verbal troubleshooting test. Presentation of the two test versions was counterbalanced. The verbal troubleshooting test is a work sample test designed to simulate a troubleshooting situation. The technician is presented with a fault and is asked to verbally isolate and repair an equipment fault through a series of iterative action-result steps. An expert assists with testing by providing the technician with results for all specified actions.

Technicians were tested and tutored individually. Technicians began by completing a pretest verbal troubleshooting test. Immediately after test completion, the expert test administrator made an holistic assessment of the technician's troubleshooting performance by assigning the technician a number from one to six. This score was used in matching and assigning technicians to the tutor groups.

During the tutoring period, technicians completed eleven troubleshooting problems on their assigned tutor, troubleshooting one problem a day in addition to completing their regular job duties. The untutored control group performed their regular job duties during the tutoring phase. After the tutoring period, technicians were posttested, using the alternate version of the verbal troubleshooting test.

Results

Troubleshooting Scores

Two subject matter experts independently scored the technicians' verbal troubleshooting protocols, using a modified Q-sort procedure. These protocols contained the actions verbalized by the individual technicians, along with the corresponding results of those actions. The scores awarded by the two experts were significantly correlated, r (49) = .960, $p < .05$ for Problem A and r (49) = .965, $p < .05$ for Problem B. Thus, a Problem A and a Problem B performance score was created for each technician by averaging the troubleshooting scores given by the two experts. These scores were then standardized, using the mean and standard deviation from the pretest data for each problem.

Pretest Performance

A multiple regression analysis was conducted on the pretest data by regressing pretest verbal troubleshooting score onto tutor group (Hydrive, F-15 Pneudraulics, Control), problem completed at pretest (Problem A or B), and the interaction between tutor group and problem completed at pretest. No significant interaction or main effects were observed. The full model accounted for only 5% of the variance in pretest verbal troubleshooting performance, F (5, 45) = .455. These results show that pretest verbal troubleshooting performance was comparable for technicians assigned to the different tutor groups. Furthermore, performance on the two verbal troubleshooting problems was similar, indicating that the problems are comparably difficult at pretest.

Pretest to Posttest Differences

Paired samples t-tests were calculated to determine if technicians' troubleshooting knowledge changed significantly from pre- to posttest. The results of these analyses show that only technicians tutored by Hydrive performed significantly better on the verbal troubleshooting posttest, t (19) = 4.14, $p < .001$ (see Figure 1). Verbal troubleshooting performance did not change significantly for either the technicians tutored by the F-15 Pneudraulics Tutor or for the control technicians.

Posttest Performance

The effect of tutor group and problem type on posttest performance was assessed by regressing the posttest verbal troubleshooting scores onto tutor group (Hydrive, F-15 Pneudraulics, Control), problem completed at posttest (Problem A or B), and the interaction between tutor group and posttest problem. The results of the overall model were significant, F (5, 45) = 3.91, $p = .005$, and the interaction between tutor group and problem type added incremental validity, F (2, 45) = 4.65, $p = .02$, indicating that the tutor effect varied with the two posttest problems.

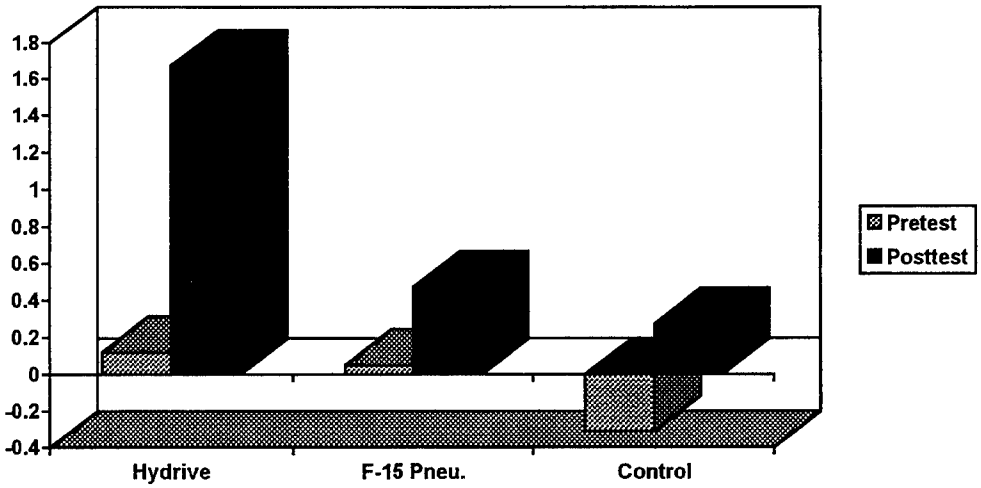


Figure 1. Average pre- and posttest verbal troubleshooting performance by technicians in the three tutor groups: Hydrive, F-15 Pneudraulics Tutor, and the no-tutor control group.

Tests of simple main effects showed that the effects of the tutor were significant with Problem A, F (2, 45) =8.75, $p < .025$, but not with Problem B, F (2, 45) = .207. Technicians who used Hydrive performed significantly better on Problem A than either the technicians who used the F-15 Pneudraulics Tutor, F (1, 45) = 5.96, $p < .025$, or the control group technicians, F (1, 45) = 7.54, $p < .01$. Technicians using the F-15 Pneudraulics Tutor, on the other hand, performed at the same level on Problem A as control group technicians who received only on-the-job training, F (1, 45) = .474 (See Figure 2).

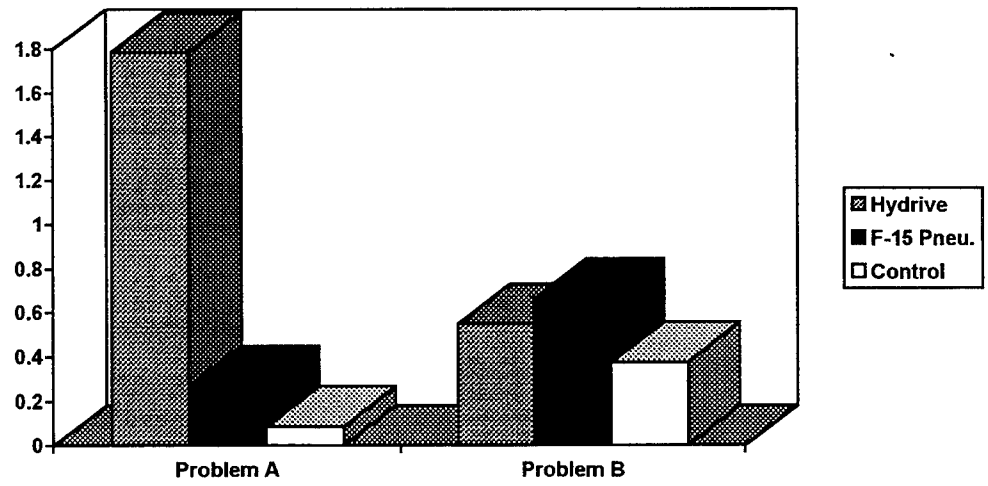


Figure 2. Average posttest verbal troubleshooting performance on Problem A and Problem B by technicians in the three tutor groups: Hydrive, F-15 Pneudraulics Tutor, and the no-tutor control group.

Discussion

The results of this evaluation indicate that tutoring on Hydrive resulted in superior performance on certain kinds of troubleshooting problems. A close examination of the two verbal troubleshooting problems suggests that this result may be related to Hydrive's emphasis on strategic knowledge: Problem A, the problem on which Hydrive technicians excelled, involves a strong strategic component because there was no fault isolation tree available for this problem, thus requiring technicians to develop their own strategy to isolate the fault. Students trained on the F-15 Pneudraulics Tutor apparently did not develop the kind of conceptual understanding that would have allowed them to solve this problem on their own. Problem B, on the other hand, involves a strong procedural component: a fault isolation tree was available that enabled technicians to follow step-by-step instructions for isolating the fault. Not surprisingly, all three groups of technicians performed comparably on this problem. Thus, the emphasis on procedural advice in the F-15 Pneudraulics tutor did not appear to provide any particular advantage to students in that group, since both groups had access to the fault isolation guides which provided the procedural steps to solve the problem.

The distinction between the strategic advice emphasized in Hydrive and the procedural advice emphasized in the F-15 Pneudraulics Tutor can be understood in terms of the system perspective on which the advice is based. That is, Hydrive's strategic advice is based on a functional breakdown of the system in terms of electrical, hydraulic, and mechanical paths and encourages space-splitting between these systems. The F-15 Pneudraulics Tutor focuses on individual components that make up the specific path in which the fault is located and focuses its advice on which components in the path have been eliminated by previous tests and which have not.

The conclusion that Hydrive's effect was attributable to the strategic feedback, as opposed to the instruction relating to system knowledge, is supported by an analysis of the records of students' tutoring sessions: In Hydrive, it is possible to obtain advice in two ways. One is by requesting it (referred to here as instruction), and the other is through intervention by the tutor (feedback). While instruction related to system functioning must be accessed by the student, strategic advice can be obtained either by the student requesting it, or through the intervention of the tutor when the student takes an action that suggests they are using an inefficient strategy. Records of Hydrive students' tutoring sessions show that they rarely, if ever, requested instruction whether it was instruction on strategies or instruction on the system. Thus, the vast majority of advice received by Hydrive students was feedback on the inefficient strategies being used, and what strategies they should try.

An alternative interpretation of the observed effect of Hydrive is suggested by examining the tutoring records for students who used the F-15 Pneudraulics Tutor. Unlike Hydrive students, these students often requested advice, with the most frequently requested type of advice being "Parts left to test." In fact, observers of the tutoring sessions commented that many of the students in this group would access this type of advice immediately upon receiving a problem and then simply swap out each part listed until they had solved the problem. This observation is supported by the large difference in time spent on each problem by students in the two tutor groups, with Hydrive students spending significantly more time than the other students. Thus,

time spent on the tutor, irrespective of the types of advice students received could account for the observed effect. According to this interpretation the fact that Hydrive students did not have available the sorts of procedural advice that enabled the other students to speed through the problems, actually enhanced its effectiveness.

It is not inconceivable that troubleshooting practice alone (without supporting instruction and advice) could provide extremely valuable learning opportunities for technicians in these domains. Given the costs to develop systems like those tested here, future research needs to systematically examine the instructional benefits of all the various capabilities provided by intelligent tutoring systems.

References

Gott, S.P. (1989). Apprenticeship instruction for real-world tasks: The coordination of procedures, mental models, and strategies. In E.Z. Rothkopf (Ed.), Review of Research in Education, Vol. 15, (pp. 97-169). Washington DC: American Educational Research Association.

Gitomer, D., Steinberg. L.S., & Mislevy R.J. (in press). Diagnostic assessment of troubleshooting skill in an intelligent tutoring system. To appear in P. Nichols, S. Chipman, and S. Brennan (Eds.), Cognitively Diagnostic Assessment. Hillsdale, NJ: Lawrence Erlbaum.

Steinberg, L.S., & Gitomer, D.H. (1994, April). Intelligent tutoring and assessment built on an understanding of a technical problem-solving task. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Fostering Students' Motivation in the College Classroom: The Role of Critical Professor Behaviors

Ann M. Herd, Ph.D.
Capt Lisa Boyce
Col Randy Stiles, Ph.D.
C1C Charlie Law
United States Air Force Academy

Abstract

Students' reports of effective and ineffective professor behaviors were investigated in relation to dimensions suggested by instructional design researchers as important classroom environment components. Data for the study were provided by 109 junior and senior-level cadets at the United States Air Force Academy enrolled in a Leadership course. Each cadet reported incidents of either effective or ineffective professor performance using an adapted critical incident technique (Flanagan, 1954). Results suggested the majority of incidents related to the classroom dimension of Personalism, or the degree to which the classroom culture is characterized by mutual respect and collaboration. Implications of these findings for instructional design are discussed.

An increasingly recognized goal of learning institutions across the country is to not only develop students intellectually but to encourage in students a learning motivation and intellectual curiosity which will stay with them throughout their lives. One such institution with an emphasis on this goal is the United States Air Force Academy (USAFA). One of the formal educational outcomes for USAFA graduates is to develop "officers who are intellectually curious". This objective is described as follows:

> Besides possessing the knowledge and having abilities to put that knowledge to use, graduates of the Academy must be inclined to do so. We want to develop an attitude of intellectual curiosity in our graduates that predisposes them to lifelong learning.

Developmental Instruction Model

Educational researchers have suggested many important instructional design issues for consideration in striving to foster students' motivation and intellectual curiosity. A developmental instruction model is based on the premise that learner development depends on the correct assessment of important learner characteristics (such as knowledge, cognitive skills, attitudes toward learning) and the interaction of these characteristics with teacher characteristics and classroom environment characteristics. According to classic work by William Perry (1970) on college students' intellectual and ethical development, the ideal college classroom environment provides appropriate amounts of both challenge and support for the student, based on the correct

assessment of the students' important learner characteristics. Specific classroom environmental dimensions proposed by researchers as important include structure, diversity, experiential learning, and personalism (Knefelkamp, 1981; Widick, Knefelkamp & Parker, 1975). These dimensions are further described below.

According to Knefelkamp (1981), structure refers to the amount of direction and guidelines provided to students regarding the course and its parameters. Examples of activities which provide structure include providing specific objectives for the overall course and each lesson, providing outlines and notes of course material, providing explicit criteria and examples for grading, providing practice opportunities (e.g., quizzes and exercises) and specific feedback on students' performance.

The environmental dimension of diversity refers to depth versus breadth of material presented in the course, or the complexity versus quantity in the number of alternative perspectives presented and studied. A common complaint of many college instructors is the felt need they experience to "get through" all the material outlined in the curriculum handbook for their course. Many college courses are designed to cover a large breadth or quantity of material, with the resulting effect that time does not allow for much depth in coverage.

The environmental dimension of experiential learning refers to the degree to which students are involved in activities which provide direct and concrete examples of course principles. Examples of activities which promote experiential learning include case studies, role play exercises, and group tasks. Asking practical questions and students' opinions, as well as providing "real world" examples of course concepts, are also ways to increase students' experiential learning.

The fourth classroom environment dimension, personalism, refers to the degree to which the classroom environment promotes a culture of mutual respect, responsibility, and collegiality. A classroom culture with a high degree of personalism is one where the instructor exhibits enthusiasm, empathy, and sincere concern for students' learning. The students in turn perceive a non-punishing environment where they are collaborative participants in the learning process.

Objectives of the Present Study

While a variety of educational researchers from varying perspectives have agreed that the instructional design dimensions reviewed above are important when designing a developmental classroom environment, few studies have investigated learners' perceptions of the importance of these dimensions. The objective of the present study was to investigate students' own reports of specific instructor behaviors which they perceived as particularly effective or ineffective. These student reports could then be content analyzed in light of the four classroom environment dimensions described above, to determine the frequency with which each of the dimensions is reported by students. The reported behaviors in each dimension could then be viewed as one indication of potential motivating behaviors as perceived by USAFA students.

Method

Participants

Participants in the study were 109 USAFA volunteer cadets from seven sections of the Behavioral Science 310 course in Leadership. Students from this course were chosen as participants because they are juniors and seniors who come from a cross-section of all USAFA majors and thus could be expected to have a variety of classroom experiences. In any given semester, approximately 60% of students in the Leadership course are Humanities and Social Sciences majors, while 40% are Engineering and Basic Science majors.

Questionnaire

Participants were asked to complete a Critical Incident Form, adapted from the critical incident technique of job analysis proposed by Flanagan (1954) (Bernardin & Russell, 1992). The form instructions directed the student to recall noteworthy examples of professor behaviors that illustrated either unusually effective or ineffective performance. The student was to choose one example and write about it by answering three questions: "1) What were the circumstances leading up to this example? 2) What, exactly, did the professor do? Describe exactly what was done (the professor's behaviors) that qualifies the example as either effective or ineffective. 3) What were the results or outcomes of the actions?"

Procedure

Cadets' Critical Incident Forms were content analyzed by three raters, who independently categorized each behavior reported on the form into one of the four categories discussed above. Mean interrater reliability among the three raters was approximately 85% (.92, .84, and .78 pairwise).

Results

Of the 109 incidents collected, 45% were effective examples and 55% were ineffective examples. The majority of cadets reported at least two specific, distinct professor behaviors in their example on the Critical Incident Form, yielding a total of 231 behaviors reported. The behaviors reported within each example often referred to different instructional design categories, indicating these dimensions were not perceived as independent.

Figure 1 presents the overall content analysis results of the reported behaviors. As shown in Figure 1, the majority of behaviors (53%) fell into the Personalism category. While 26% of examples referred to Experiential Learning activities and 16% of the examples reported referred to Structure behaviors, only 5% of the behavioral examples referred to the Diversity dimension.

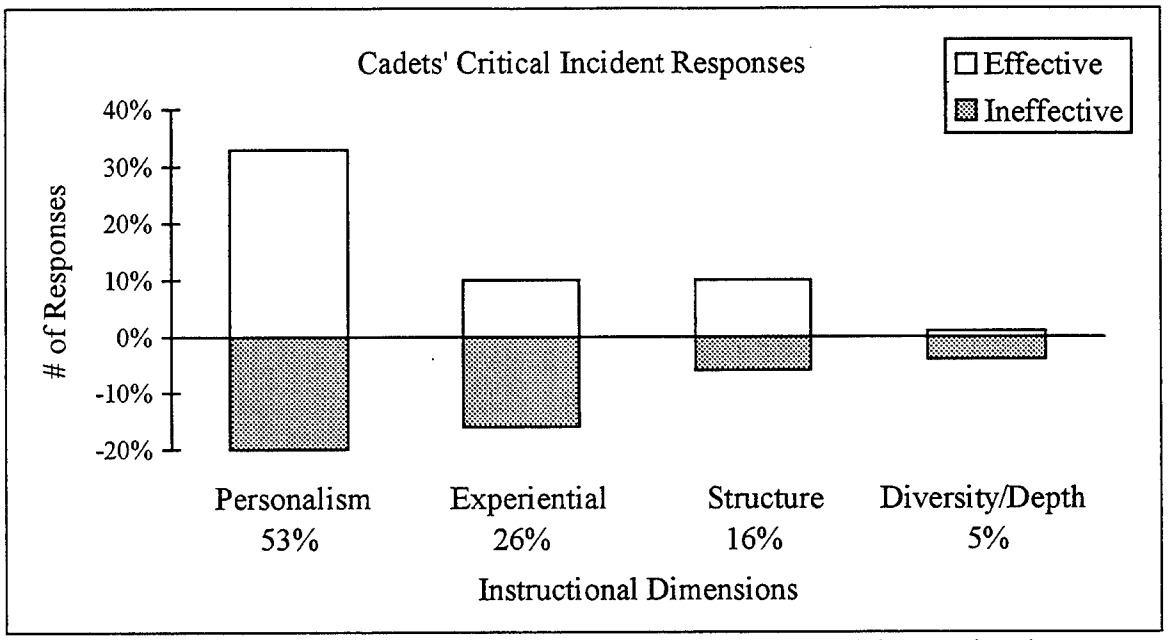


Figure 1. Percentage of reported critical professor behaviors in each instructional category.

Discussion

Results of the study suggest that USAFA cadets are more likely to generate and report examples of unusually effective or ineffective professor performance which pertain to the Personalism dimension of the classroom environment. Cadets also generated many examples of Experiential Learning but were least likely to report examples which pertained to classroom Structure and Diversity.

Explanations for the preponderance of Personalism examples include the fact that USAFA cadets work in a particularly stressful environment which requires long hours of hard work in the areas of academics, athletics, and military discipline, and which also requires strict adherence to many rules and regulations. Thus, professors who encourage a classroom environment which is friendly, nonthreatening, and characterized by mutual respect and collaborative learning may be especially appreciated by cadets. Likewise, cadets report a better learning environment is characterized by opportunities for Experiential Learning, where they can actively participate in more concrete and interactive learning activities. It should be noted that in both of these categories, students reported both positive and negative examples. Thus, classroom learning experiences characterized by high degrees of Personalism and concrete experiences in Experiential Learning are considered particularly effective by cadets, while classroom experiences characterized by low degrees of these dimensions are recalled as particularly ineffective.

The critical incident procedure used in the study is both a strength and potential limitation. Advantages of the critical incident technique used in the present study include the fact that students were not "led" by the instructions to give examples pertaining to any particular instructional design dimension. Cadets were simply asked to report an example of either effective or ineffective professor performance. Because the instructions focused on professor behaviors, the resulting examples may simply reflect the dimensions that students perceive as more under the

direct control of their professors, perhaps explaining the lower focus of student comments in the Structure and Diversity dimensions. Although it could easily be argued that all four instructional design dimensions studied can be directly influenced by professor behaviors and choices in the classroom, future studies may benefit from asking students directly about these specific dimensions.

Implications of the study findings include specific suggestions regarding professor behaviors which are perceived by students as particularly effective or ineffective for their learning. A USAFA instructor interested in fostering student's motivation in the classroom by increasing students' perceptions that their classroom environment exhibits a high degree of Personalism might, for example, learn and use each student's name, get to class early and talk to students, use humor in the classroom, show genuine concern and give assistance when students do not understand course material, and role model the behaviors expected from students. Punishment techniques and displays of anger were reported by students in this study primarily as negative examples of Personalism. Thus, it seems likely that punishment and anger are effective only when students know their learning is the professor's primary concern.

In summary, the present study suggested that students at the United States Air Force Academy perceive professor behaviors fostering a collaborative and interactive environment as important for their learning. Future studies are needed which directly assess the relative perceived importance and interaction of instructional design variables. Future investigations should also assess the learner characteristics in a variety of settings which influence these perceptions.

References

Bernardin, H.J. & Russell, J.E.A. (1992). Human resource management: An experiential approach. New York, NY: McGraw-Hill, Inc.

Flanagan, J.C. (1954). The critical incident technique. Psychological Bulletin, 51, 327-358.

Knefelkamp, L.L. (1981) Developmental Instruction. Counseling and Personal Service Department: University of Maryland.

Perry. W.O. (1970). Intellectual and ethical development in the college years: A scheme. New York: Holt, Rinehart & Winston.

Widick, C., Knefelkamp, L., & Parker (1975). The counselor as a developmental instructor. Journal of Counselor Education and Supervision, 14, 286-296.

Two Internal Yardsticks for Integrity

Colonel Clark Hosmer, Ph.D.
Shalimar, Florida

Abstract

Each of the five United States academies has an honor system. Although the honor systems differ in detail, each fosters personal integrity. Nevertheless, from time to time cases of personal dishonesty occur. In addition to the honor systems and scores of external moral guidelines of society, this paper offers two yardsticks that would be internal, inside the student's head, First, go by long-term consequences. Second, assume your decision will be publicly known. An implication of the conclusion is that more work on the yardsticks for a broader topic may be justified.

The goal is to have every graduate of the five taxpayer-funded, United States academies serve the nation with integrity. From time to time, however, the academies find a student has committed a dishonest act. My purpose is to offer two yardsticks for integrity. Use of the yardsticks could help academy students achieve a "standard of honesty and moral strength" cited by Lt. General Paul E. Stein, Superintendent of USAFA in his Instruction 63-158. General Ronald R. Fogleman, The Air Force Chief of Staff, in the Policy Letter Digest of the Office of the Secretary of the Air Force said, "Because of what we do, our standards must be higher than those of society at large."

Many Guidelines for Integrity

The cadets and midshipmen in the United States Academies have many guidelines for integrity. The academies' honor systems and codes are central. Also, before entering an academy, each young man and woman has built a personal world of guidelines for morality from home, school, church, and peers on the street. Moral guidelines that are cited by Bartlett's Familiar Quotations include from 2400 B.C. in Egypt, "Truth is great and its effectiveness endures," 15 Greek and Roman, eight Oriental, and 11 Bible pronouncements on truth, plus statements by modern philosophers such as Alfred North Whitehead. He wrote, "There are no whole truths; all truths are half-truths. It's trying to treat them as whole truths that plays the devil," Perhaps the most abundant guidelines are moral admonitions that come from pulpits. Religions have the reputation of being a primary source of guidelines for morality.

Cadets and midshipmen are carefully selected. They are bright and know that honesty has compelling payoff. But they also know the world of high school cheaters. "Cheating in Our Schools: A National Scandal." by Daniel Levine (1995), reported that a national survey of 3100 top students found eight out of ten say that they cheat. In contrast, in one of the major three U. S. academies, over a period of five years the average number of individual cases of dishonesty of all

categories was less than 6% per year. The difference in favor of academies does not account for instances of dishonesty that are not discovered. Nevertheless, the difference suggests that the academies' honor systems do well in meeting a difficult challenge. They produce a society dramatically higher in honesty than the wholesale dishonesty found in the culture from which the students were drawn.

No one guideline for integrity satisfies everybody. Nor do all the guidelines keep everybody honest all of the time.

Testing for Right and Wrong

People's actions produce consequences that are good and bad for people: Let us define good and bad as:

A good act benefits people.

A bad act hurts people.

"People" of course includes one's self. Benefits and hurts can be physical, psychological, and economic.

Please note that something is missing. "Sinful" or "evil" acts do not play a part in the above definitions. The reason is that when we moralists label an act as sinful or evil, our label boils down to a prediction that the consequences of the act will hurt people. Therefore, all such acts are bad. To deal with only consequences keeps our focus on causal relations and avoids the intensity of crushing sin. Hammering sin tends to be a complicating overload in problem-solving.

Assume a cadet or midshipman has an opportunity to use marijuana. The question of what to do is clear: Will he go along with his anticipation of drug-affected feelings of well being and camaraderie with his friends? Or, will he anticipate as the drug phases down, his emotional let down, the risk of being caught and dismissed, and the long term adverse impact on his health?

What he decides to do depends on whether he will go by the short-term or by the long-term consequences of his decision.

The First Yardstick for Integrity

Go by the long-term consequences of the act..

The yardstick is an approximation. First, to begin with prediction is chancy. The long-term consequences surely are more complex than indicated in this brief. Second, the situation may require a quick decision. The feeling of pressure to solve the problem of discounting the short-term and working his estimate of the long-term consequences, may founder the cadet or midshipman. What the situation calls for is the "moral strength" cited by General Stein. Let us see if one of the standard moral guidelines would help..

Take conscience. A typical conscience consists of conformance to the accepted moral values of the youth's community, His conscience depends on the lessons he has learned . If he is from a family of a church leader, he might enter his teen years with a conviction that sinful temptations abound. He may be prepared to deny himself any sampling. From the home of Mac the Knife, however, he might be tolerant of shop lifting as a means of income. Another guideline is poet George Crabbe's use of habit as a test of morality: "It must be right: I've done it from my youth," Habit may be reliable but its validity may be questionable.

The conscience as a guideline for integrity is like a corral of posts and rails to induce one from straying out to attractive but dangerous territory. The posts and rails of the corral are loosely set. A wandering eye may find ways to slip out of the corral for out-of-bounds dalliance with one or more of the temptations that abound.

A cadet's or midshipman's conscience might be as Shakespeare wrote in Richard III, "My conscience has a thousand several tongues." Similarly, Luigi Pirandello wrote "Don't you see that that blessed conscience of yours is nothing but other people inside you?" In effect, the moral strengths and the moral limitations of those other people are the cadet's and the midshipman's conscience. The possible variations within each conscience may dilute its effectivenss as a guideline for integrity.

Nevertheless, conscience plus a personally selected array from the myriad of moral guidelines are the cadets' and midshipman's armamentarium with which to handle his moral situation.

To help reinforce his moral strength teach him to go by estimated long-term consequences.

"The Second Yardstick for Integrity

H. L. Mencken wrote, " Conscience is the inner voice which warns us somebody may be looking."

If the cadet or midshipman still squirms before his decision, apply Mencken's lesson.

The second yardstick is: Assume that your decision will be published.

Conclusion

If the cadets and midshipmen at the five United States academies internalize the suggested yardsticks for integrity, the frequency of individual acts of dishonesty probably would be reduced.

The use of the two yardsticks needs to be as nearly automatic as training can bring into being. Perhaps the student honor committees could include in their training programs the usefulness of using 1) the long-term consequences of, and 2) assumed public knowledge of, personal decisions

on moral problems. The yardsticks could operate inside the heads and hearts of cadets and midshipmen as automatic aids to achieve and sustain a "standard of honesty and moral strength."

Implication

Although this paper starts and ends with focus on only the students in the five United States Academies, an implication of the conclusion is that the relevancies of the yardsticks to our culture may justify future work toward a broader topic paper.

References

Bartlett's Familiar Quotations. (1992). 16th Edition, Boston: Little, Brown & Co.

Durant, W. (1953). The Story of Philosophy, New York: Simon & Schuster.

Levine, D. R. (1995). Cheating in Our Schools: A National Scandal The Reader's Digest October.

Merriam-Webster's Collegiate Dictionary, (1994). Tenth Edition, Springfield: Merriam-Webster.

Policy Letter Digest. (1995). Office of the Secretary of the Air Force, Washington, DC: December.

United States Air Force Instruction, (1995). 36-158, United States Air Force Academy. August.

Gender and Scholastic Aptitude Test Scores: Relationship to Grade Point Averages at the United States Air Force Academy

Dawn L. McCown
Justin D. Rueb
United States Air Force Academy

Abstract

Conflict arose at the beginning of the equal rights movement about the equality of females to males in the area of cognitive ability. Studies such as Anastasi (1958) and Tyler (1965), support the idea of gender differences, whileothers, like Hyde and Linn (1988), or Kaufman (1988) show nosignificant differences in gender cognitive skills. Traditional beliefs suggest males are superior at math related skills, while females are superior in verbalassociated tests. Some typical tests have supported the traditional differences between males and females. Scholastic Aptitude Test scores have shown significant score differences between males and females at the Air Force Academy. The relative inequality on placement tests would suggest that a person's sex would affect their GPA, but this is not the case. At the United States Air Force Academy, male and female cadets perform similarly in academics according to grade point averages.

"I'm not denyin' the women are foolish; God almighty made 'em to match the men" (Elliot, cited in Bartlett, 1982). This quote, by George Elliot, exemplifies the spirit of modern day America. The United States is becoming a land of increasing equal opportunity where women and men are competing for the same jobs, based on performance, not gender. Although equality is an ideal many citizens try to live by, many questions have been raised about intellectual equality between the sexes. Do women have the same intellectual ability as men? Since the feminist movement of the mid-twentieth century, many scientists (Anastasi, 1958; Naglieri, 1993), have investigated gender differences in intelligence. The topic of academic ability between the genders is controversial, and many studies have produced conflicting conclusions.

Many early studies, (Anastasi, 1958; Tyler, 1965; Maccoby, 1974) supported the idea that cognitive differences exist. Other studies, (Hyde & Linn, 1988; Hyde, Fennema, & Lamon, 1990; Kaufman, 1988) did not find any relationship between sex and intelligence factors. If the controversy involving gender differences supports that cognitive differences do exist, than equal opportunity education may need to be specialized for each gender according to their specific talents.

An idea that gained support in the late 1960's is that men were superior to women in mathematical abilities, while women were superior in the area of verbal ability. Many studies (Anastasi, 1958; Tyler, 1965; Maccoby, 1974) have shown a relationship between sex and intelligence, but few have been considered valid, because of criticism of their test methods. The reason for the lack of an extensive knowledge base in this area is the inconsistent metric systems

used when measuring intelligence earlier in the century. Many of the tests were biased in that they favored certain genders or ethnic groups (Wolman, 1981).

Anastasi (1958) did a study on cognitive differences using measures of intelligence (Anastasi, 1958). He examined a sample of male and female subjects of varying ages through systematic intelligence testing. Anastasi concluded that boys outperformed girls on aspects of spatial ability, arithmetical reasoning, and general information. The experimenter also concluded that girls outperformed boys on tasks of verbal ability, spelling, rote memory, and perceptual speed (Anastasi, 1958). Anastasi's study generated an interest in cognitive roles that led to further studies in the area of sex roles. The new paradigm involved defining roles for the sexes assuming men and women had different cognitive talents. These test findings indicated a difference in cognitive ability, but later studies (Hyde, 1988, 1990) have challenged the ideas set forth by the pioneers of sex studies.

Challengers to the original findings of cognitive differences were Hyde and her associates (Hyde & Linn, 1988; Hyde, Fennema, & Lamon, 1990). She conducted two separate meta-analysis studies: one assessing gender differences in mathematics performance, and the other assessing gender differences in verbal ability. On the first test, Hyde and Linn (1988) investigated gender differences in verbal ability. They completed a meta-analysis of 165 different studies to test for verbal superiority in females. Hyde and Linn found gender effects for verbal ability non-existent. Hyde, Fennema, and Lamon's (1990) meta-analysis of mathematics showed nearly the same results as the verbal meta-analysis. They concluded that gender differences were near negligible, with the greatest difference between males and females being in the area of problem-solving tasks. The Hyde studies opposed the findings of the earlier scientists, and helped support the idea of cognitive equality between the sexes.

When studying cognitive ability differences between the sexes, other variables need to be considered because they could significantly alter the results of the study. The first variable is age. Depending on what ages are studied, the outcomes could be different. Many tests disagreed about age effects on the differences in cognitive abilities between males and females. Warrick and Naglieri (1993) and Benbow and Stanley (1980), both showed age had an extensive effect on gender specific cognitive ability. Earlier studies (Anastasi, 1958; Tyler, 1965) failed to account for age differences in their experiments. Warrick and Naglieri completed a study where they tested verbal and mathematical sex differences at grade 3, grade 6, and grade 9. Their findings indicated third grade girls were significantly better at attentional verbal processes than boys, yet no significant differences existed for the sixth and ninth grade samples. The results of their tests suggested female superiority in verbal skills occurred at a young age. Female superiority diminished in the teens and equality between the sexes in verbal ability was carried on through adulthood.

In contrast, Benbow and Stanley (1980) compared age with mathematical ability by analyzing the data gathered by the Study of Mathematically Precocious Youth. They found that males began to show superiority to females in mathematical ability as they got older. Up until the seventh grade, mathematical abilities showed little differences between the sexes, but after the seventh grade, a significant difference between the mathematical abilities of boys and girls existed.

The greatest noticeable difference was in the upper ranges of mathematical reasoning ability. The results suggested males would retain their mathematical superiority into adulthood, while females would lose their verbal advantage before reaching maturity.

Environmental and societal factors may also contribute to male superiority in mathematics. These variables can be used to explain why female mathematical ability decreases with age. By looking strictly at the variable of society, the decrease in female mathematical performance could occur because society has traditionally encouraged boys to pursue mathematically based education. Benbow and Stanley (1980) found females take calculus in high school 35% less than their male counterparts. Female deficiency in calculus courses could be due to our society, or to women having increased cognitive difficulties in mathematics as they mature.

A second important aspect in assessing grade point average is to analyze college entrance examination scores. If tests like the SAT are good predictors of which students should be admitted to college, then grade point averages should be higher for students that receive high SAT scores. If academic ability is the same for males and females, SAT scores should be comparable, and thus GPA's among the sexes should be similar, also. However, SAT scores are not always found to be comparable between the sexes.

Benbow and Stanley (1980) found adolescent men and women received different scores on the SAT-Q and SAT-V tests, respectively. The SAT-Q is an assessment of mathematical ability, while the SAT-V test is an assessment of verbal ability. On SAT-Q tests, adolescent boys did significantly better than adolescent girls. The 1972 SAT-Q test given to gifted students chosen for a talent search reported significant results. Out of the entire sample, 27.1 percent of the boys received a score over 600, while none of the girls received a score over 600. Although criticized for their methods, the findings by Benbow and Stanley were significant and provide support that cognitive differences exist between the genders.

Later, Hyde, Fennema, and Lamon (1990) didn't find the SAT-Q to elicit such surprising results. In fact, they found the SAT-Q to show a very small relationship between sex and SAT-Q scores that favored males, but not one that was as strong as the Benbow and Stanley findings. The study did show a significant difference though, even if only a small one. The Hyde, Fennema and Lamon (1990) study would lead us to believe that on average, males score better on the math portion of the SAT than females.

Hyde and Linn's study on the verbal portion of the SAT didn't seem to elicit any differences between the sexes. Women have not scored significantly better on the SAT-V since 1972 (Hyde & Linn, 1988). Because women have not shown superiority in verbal tasks, the distribution of scores is about the same for males and females. The relative equality between males and females in verbal tasks would lead us to believe that scores between the sexes would be comparable on the SAT-V as well as GPA in verbal subjects. By looking at the variables of cognitive sex differences and SAT scores, I believe that male cadets will have higher grade point averages than female cadets due to the math based curriculum present at the United States Air Force Academy.

Method

Sample

Three hundred subjects were randomly selected from an academic database of juniors and seniors at the United States Air Force Academy. For each subject, class year, gender, GPA, SAT-V and SAT-Q scores were generated. These classes were chosen because their cumulative GPAs are well established, giving a more accurate description of the cadet's academic potential. Two hundred and fifty males and fifty females comprised the sample, which is roughly representative of the percentages at the Air Force Academy. The mean age was 21.33 years.

Measures

Each subject's grade point average was used to assess their academic success. Grade point averages at the Academy are on a 4.0 scale. Cadets in academic trouble (GPA < 2.0) are relatively few since Academy standards for maintaining a minimum grade point average are strict. Cadets with a grade point averages less than 2.0 usually leave the Academy before the beginning of their junior year.

SAT scores are thought to be an accurate predictor of intelligence. (Wolman, 1981). Consequently, the SAT-V (the verbal portion of the test) and the SAT-Q (the mathematical portion of the test) were used to assess whether a difference in verbal or mathematical ability existed between males and females at the Academy.

Results

Pearson correlations were run to determine the relationship between SAT scores and GPA. T-tests for independent samples revealed the verbal portion of the SAT showed little difference ($p>.05$) between the verbal abilities of males and females. The mean score for males was 573 ($SD = 59.39$), with the scores ranging from 280 to 730 (Table 1). The female sample was very similar, with a mean of 575 ($SD = 66.74$), and a range from 450 to 770.

Table 1
Male and Female Performance in GPA, SAT-V, and SAT-Q

Variable	Male		Female	
	Mean	Std Dev	Mean	Std Dev
GPA	2.85	.46	2.85	.41
SAT-V	573	59.39	575	66.74
SAT-Q	665	71.58	637	61.25

The math portion of the SAT showed differences ($p<.05$) between male and female cognitive abilities in mathematics. The mean of the male sample was 668 ($SD = 57.81$), with a range of 290 (Table 1). The female sample produced a mean of 637 ($SD = 61.25$), and a range of 230.

Finally, no difference between the sexes existed for grade point averages. Male mean GPA was 2.85 (SD = .457) (Table 1), which was identical to female mean GPA (SD = .410) (Table 2) for the sample used. Again, the distributions were normal.

Discussion

The results showed a relationship between GPA and SAT scores did not exist. The differences in cognitive mathematical abilities when students enter the Academy did not affect their academic performance. Female cadets were as likely to succeed in academics as their male counterparts. Although SAT scores may be a rough predictor of success in college, they do not accurately predict success differences between males and females in mathematically based curriculums. For cadets at the Air Force Academy, high Scholastic Aptitude Test scores are needed to be accepted, but once accepted, data suggests that equal opportunity exists regardless of sex.

Although results of the study are significant, the use of cumulative grade point averages may not have been the most accurate predictor. Although the curriculum at the Air Force Academy is mathematically based, some non-technical classes are required. For example, English and foreign language are requirements for incoming freshmen. The problem of including non-technical classes in the study could have been avoided by using GPAs corresponding solely to mathematically based classes. Another problem with using cumulative grade point averages is that major's classes are included in the statistic. Typically, by the time cadets are juniors, they have begun their major's classes, after taking two years of basic core classes. Because the cadets used in the sample were juniors and seniors, major's classes performance was included in the grade point averages used. Again, ideally, only core GPAs corresponding to math would be used.

Regardless of these problems, the results suggest a positive academic atmosphere for all cadets at the United States Air Force Academy. Although significant difference existed between male and female SAT-Q scores, the difference the abilities have on grade point averages were nonexistent. Studies in the future could continue to focus on intelligence and gender relationships. One way this could be done is by looking at the relationship between ACT scores and grade point averages. A relationship between ACT scores and GPA could show that the ACT is a better predictor of success than the SAT. Another idea could be to do a study that includes many universities using only mathematical grade point averages. A wider sample of SAT scores could be obtained, eliminating the possibility of the Academy selection process bias.

This study found no relationship between gender differences in grade point averages for cadets at the United States Air Force Academy. The study suggests less emphasis should be put upon SAT scores in choosing college freshmen in the future. Colleges and Universities in the recent past have started to take other factors besides entrance examinations into account. The emphasis seems to be on leadership and athletic activities as opposed to testing scores. This trend for a more well rounded student has occurred at the Air Force Academy, as well as other institutions of higher learning. Colleges and universities should continue on this trend, because these measures may be more accurate in predicting success in college than entrance examination scores. De-emphasizing entrance examination scores will help to put females on a more equal

basis with their male counterparts, as well as allow colleges and universities to create a better product in the future.

References

Anastasi, A. (1958). Differential psychology (3rd ed.) New York: Macmillan.

Bartlett, J. (1982). Familiar Quotations. New York, NY: Little, Brown and Company.

Benbow, C. P., & Stanley, J. C. (1980). Sex differences in mathematical ability: Fact or artifact? Science, 210, 1262-1264.

Eysenck, H.J., & Kamin, L. (1981). The intelligence controversy. New York, NY: Wiley-Interscience Publications.

Feingold, A. (1993). Cognitive gender differences: A developmental perspective. Sex Roles, 29, 91-112.

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. Psychological Bulletin, 104, 53-69.

Hyde, J. S., Fennema, E. & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. Psychological Bulletin, 107, 139-153.

Kaufman, A. S., Mclean, J. E., & Reynolds, C. R. (1988). Sex, race, residence, region and education differences on 11 WAIS-R Subtests. Journal of Clinical Psychology, 44, 127-133.

Rosenthal, R. & Rubin, D. B. (1982). Further meta-analytic procedures for assessing cognitive gender differences. Journal of Educational Psychology, 74, 708-712.

Warrick, P. D., & Naglieri, J. A. (1993). Gender differences in planning, attention, simultaneous, and successive (PASS) cognitive processes. Journal of Educational Psychology, 85, 693-701.

Wolman, B.B. (1981). Handbook of intelligence: Theories, measurements, and applications. New York, NY: Wiley-Interscience Publications.

Psychological Androgyny and its Relationship to Leadership Grades of Cadets at the United States Military Academy[2]

Captain Lori Anne Stokan, M.S.
Sehchang Hah, Ph.D.
United States Military Academy

Abstract

Psychological androgyny theory (Bem, 1974) states that individuals can display appropriate sex-role characteristics across a variety of situations. An androgynous individual can be a more effective leader by being cognizant of the changing constraints of a situation and engaging the most effective behavior for that particular situation. The relationship of psychological androgyny theory to leadership was tested using five hundred and twenty-seven cadets in their first and third year at the United States Military Academy (USMA) by comparing their sex-role types to their military development grades. The data showed that cadets in their third year at the Academy were not more androgynous than cadets in their first year despite having more leadership training and experience. Analysis also showed that no significant relationship existed between the sex-role type and the cumulative military development grades. Theoretical and practical implications of the findings are discussed in terms of directions for further research and leadership training at the USMA.

Prior to the 1970's, society's cultural norms defined masculine and feminine traits as opposites (Bem, 1981). Masculinity traditionally had been, and, some will argue, still is associated with a task-oriented, directed approach that values rational problem-solving. On the opposite extreme, femininity has been characterized as a people-oriented, supportive approach that values sharing feelings and caring for others. The theory of psychological androgyny disputed these traditional sex-role stereotypes by postulating that a person can exhibit both masculine and feminine qualities and behaviors. One's effectiveness is also enhanced through the equal valuing and expression of both masculine (instrumental or task-oriented) and feminine (expressive or people-oriented) behaviors (Bem, 1974).

The theory of androgyny remains pertinent today as organizations require androgynous managers. Blanchard and Sargent (1984) argued that the effective manager of the future will be a "situation" leader who shows behaviors of both extremes, depending on the environment and the needs of the people involved. Concern for people is approaching parity with concern for getting the job done. Naisbitt (1992) echoed this concern by discussing the trend for "high-touch" people

[2]The authors wish to thank COL Patrick Toffler, USMA and Dr. Maureen Callahan at Long Island University, thesis advisor. The research is based on the masters thesis of the first author. The views expressed herein are solely the authors' and do not represent the views of the USMA.

contact that accompanies any "high-tech" advance. Contemporary leadership theory espouses that effective leadership results from a combination of interpersonal skills and task awareness--androgynous behavior (Powell, 1988). The androgynous manager is able to draw upon his or her instrumental abilities to facilitate the completion of tasks or "getting the job done," while drawing upon his or her expressive abilities to show compassion and support for employees and maintain morale (Lord, 1977).

The purpose of the United States Military Academy (USMA) is to provide the nation with leaders of characters who serve the common defense (USMA, Strategic Guidance, 1993). To that end, the staff and faculty at West Point spend much time and effort on leadership development of cadets. The underlying philosophy of leadership development at West Point is that people acquire and develop leadership through practice across a variety of situations. This practicing is supervised and assessed against an established standard with feedback given on performance. An evaluative subsystem, based largely on the grading of a cadet's leadership performance through the assignment of the military development (MD) grades, focuses primarily on the institution's need to differentiate performance among cadets (USMA, USCC Reg 623-1, 1994).

The focus of this study was to prove or disprove two hypotheses. The first hypothesis was that there would be a greater number of androgynous cadets in their third year at the USMA than cadets in their first year. By assessing the relationship between cadets' military development grades and their sex-role identity (androgynous, undifferentiated, masculine or feminine), we should be able to determine if the leadership training and experience after three years at West Point have been successful in developing an androgynous style of leadership in cadets. The second hypothesis was that cadets who are androgynous in their behavioral preferences will receive a higher military development grade than cadets who are masculine or feminine role-typed. Does the Academy's leadership acknowledge the importance of androgynous leadership by awarding higher military development grades to androgynous cadets?

Method

Participants

Seven hundred and four cadets from the approximate population of 4,000 cadets at West Point were initially asked to participate in the study. That total consisted of 100 males and 100 females from the Class of 1998 (CL 98) who were randomly chosen. Additionally, 52 females and 452 males from the Class of 1996 (CL 96), the total population of junior-year cadets who had held the position of squad leader as a summer assignment, were also asked to participate on a voluntary basis. The end sample consisted of 527 cadets: 72 males CL 98, 77 females CL 98, 338 males CL 96, and 40 females CL 96.

Procedure

Participants answered the long form of the Bem Sex-Role Inventory (BSRI), a self-report measure composed of 60 adjectives: twenty socially desirable masculine characteristics, twenty

socially desirable feminine characteristics, and twenty neutral characteristics. The participants rated themselves on a scale from "1" (never true) to "7" (almost always true) on each characteristic.

An androgyny score was derived by a median split procedure (Bem, 1981). The median raw-masculinity (RAW-Masc) and raw-femininity (RAW-Fem) scores were determined for the whole sample and participants were classified with respect to the respective group median as part of one of four groups representing the individual's sex-role type: androgynous (RAW-Masc and RAW-Fem scores were equal to or exceeded the medians), undifferentiated (RAW-Masc and RAW-Fem scores were both lower than the medians), masculine (RAW-Masc was equal to or exceeded the RAW-Masc median, RAW-Fem was lower than the RAW-Fem median), and feminine (RAW-Masc was lower than the RAW-Masc median, RAW-Fem was equal to or exceeded the RAW-Fem median).

Results

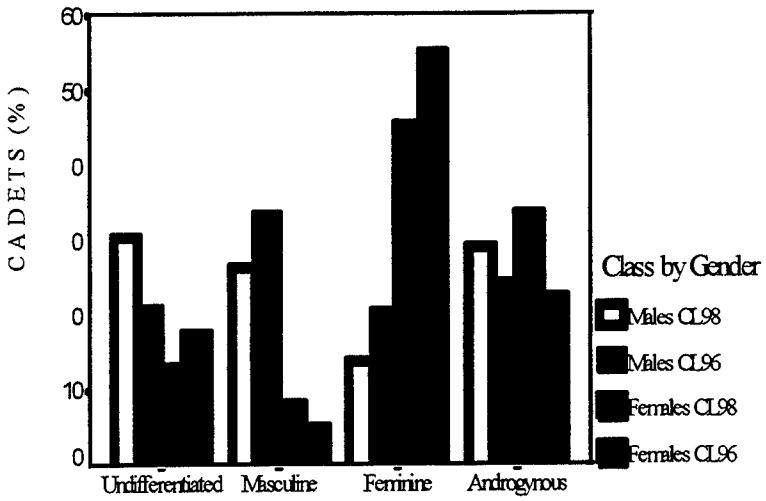
For both male and female cadets, the proportions of androgynous cadets were larger in CL 98 than in the CL 96 (Figure 1). This is contradictory to the first hypothesis since cadets were inclined to be less androgynous with more training at West Point. It also appears as though the longer male cadets stayed at West Point, the more masculine or feminine role-typed they became. Female cadets appeared to be more feminine sex-typed in their third year. They did not appear to take on the masculine role-typed behaviors of the institution which maintains a 90:10 ratio of women to men.

Figure 1. Percentage of cadets in sex-role types

As shown in Figure 2, the military development scores are spread evenly among the four sex-role type categories. The figure shows no trend of higher or lower military development scores in any of the sex-role types. Analysis of variance (ANOVA) was used to test the second hypothesis that androgynous cadets would receive higher military development scores than masculine or feminine role-typed cadets. The test showed that there was no significant difference among sex-role types on military development scores after factoring out the variance of sex and year-group variables, $F(3, 521) = 0.486$, $p > 0.05$.

Discussion

The results of the present study indicated that the first hypothesis was not supported. Despite a greater range of leadership experiences and training to draw upon, cadets in their third year were not more androgynous, but rather showed the tendency of being more sex-typed than the first year cadets (Figure 1). If officials espouse that an effective leader must have a wide range of skills, leadership classes could address the issue of using both people- and task-oriented approaches as the situation dictates. By learning a greater repertoire of behaviors, cadets would be better equipped to face a myriad of leadership challenges.



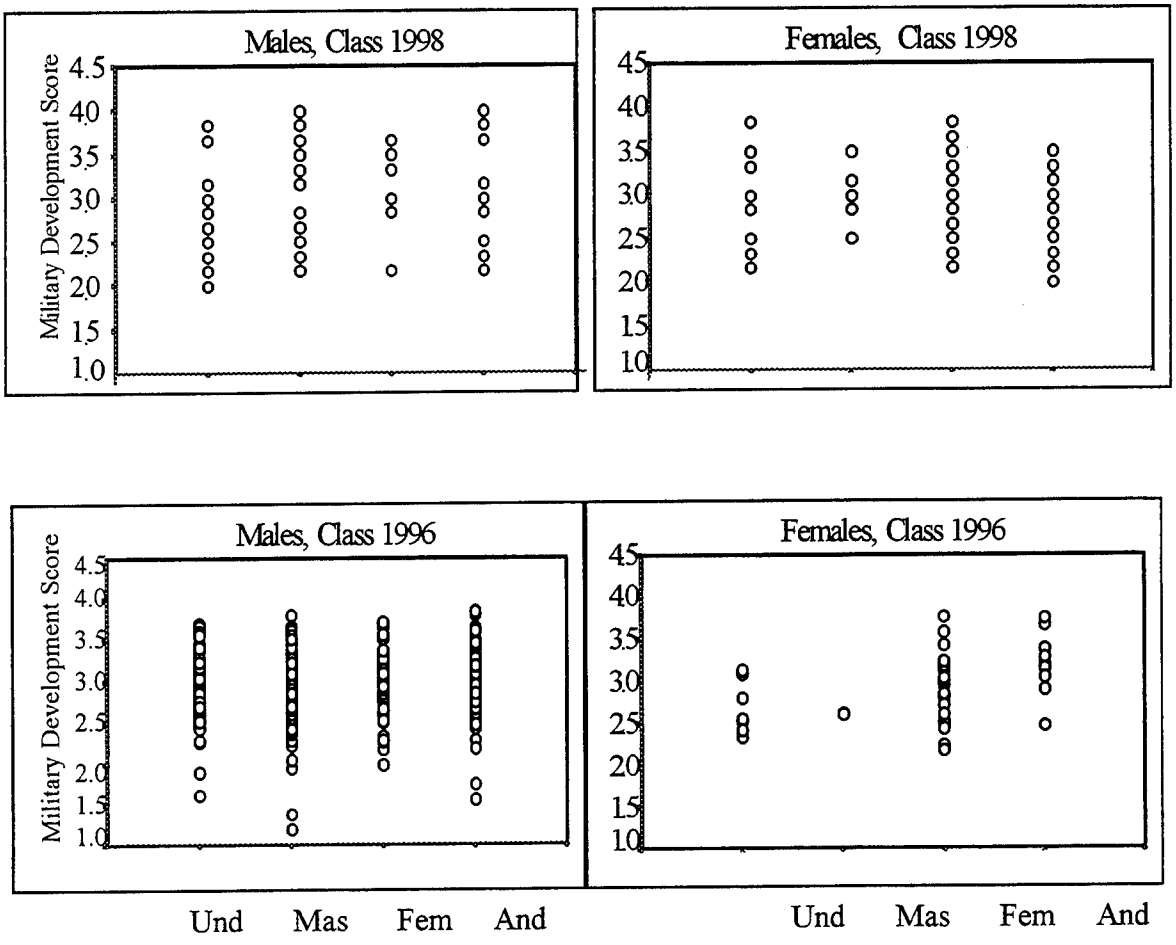


Figure 2. Military development scores across sex-role types sorted on class and sex. (Note: Und = Undifferentiated, Mas = Masculine, Fem = Feminine, And = Androgynous; grades were converted from scores on a 4.0 scale.)

The second hypothesis was not supported either. No significant relationship was found between military development grades and sex-role types. This finding was surprising since contemporary leadership theories propagate the use of both task-oriented and relationship-oriented behaviors. The results may indicate that androgynous cadets are not being recognized for their greater repertoire of behaviors. While leadership theories such as the situational leadership theory are being taught to cadets at West Point, these theories may not be actualized as more effective in practice, or rewarded as such. West Point officials may not be incorporating both instrumental and expressive traits in their leadership training. Another possibility is that while the leadership theories are taught in an academic environment, officers and cadets may not be using these theories in practice. They may not perceive these traits as related to leadership. Lastly, the items of BSRI may not adequately describe traits of a leader.

This study has many implications for further study. Future research could focus on a longitudinal study of the same cadets to see if the trend does indicate a more androgynous outlook after a greater amount of leadership experience and training. As the cadets progress into their Army career, they may become more androgynous. Similar research could be conducted on what traits leaders in the Army view as necessary for a good leader. By asking officers at West Point to determine the likelihood of each of the expressive and task-oriented traits on the BSRI for a good leader, we could determine what Army leadership espouses as traits that a leader should have. These traits could indicate a more androgynous leader as the model leader in today's Army.

Future research also could determine if a bias in response patterns actually exists. Instead of using the preconceived masculine and feminine traits on the BSRI, we could use traits which are not sex-specific (e.g. ninety percent of women might say they would be very likely to be compassionate, while only forty percent of men might say they would be.) West Point officials could then determine which traits needed reinforcement. Cadets would be trained to know, for instance, that in order for a good leader to show consideration for others, a bedrock value of West Point, he or she must use the trait of compassion. Academy leaders could encourage a wider range of responses through leadership education and training.

References

Bem, S. L. (1974). The measurement of psychological androgyny. Journal of Consulting and Clinical Psychology, 42, 155-162.

Bem, S. L. (1981). Bem Sex-Role Inventory manual. Palo Alto, CA.: Consulting Psychologists Press.

Blanchard, K. H. & Sargent, A. G. (1984). The one minute manager is an androgynous manager. Training and Developmental Journal, 38(5), 83-85.

Lord, R.G. (1977). Functional leadership behavior: measurement and relation to social power and leadership perceptions. Administrative Science Quarterly, 22, 114-133.

Naisbitt, J., Jr. (1992). Megatrends. New York: Warner Books.

Powell, G.N. (1988). Women and men in management. Beverly Hills, CA: Sage.

Stokan, L.A. (1995). Psychological androgyny and its relationship to leadership at the United States Military Academy. Unpublished masters thesis. Long Island University, NY.

United States Military Academy. (1994). United States Corps of Cadets (USCC) Regulation 623-1. West Point, NY: USMA.

United States Military Academy. (1994). 2002--a roadmap to our third century. (3rd ed.) West Point, NY: USMA.

An Innovative Approach to Curriculum Evaluation in a Civil Engineering Domain Panel Session

Theodore A. Lamb, Ph.D., USAF Armstrong Lab USAFA/OL
Winston R. Bennett, Jr., Ph.D., USAF Armstrong Lab/HRT
Kent L. Gustafson, Ph.D., Univ of Georgia
Kurt C. Kraiger, Ph.D., Univ of Colorado/Denver
Capt Mike Rits, USAFA/DFCE

Abstract

This panel addresses work being done by the Armstrong Laboratory Remote Operating Location at the Academy (collocated in the USAFA Center for Educational Excellence), the USAFA Department of Civil Engineering, the University of Colorado/Denver and the University of Georgia. The AL is assisting the Department of Civil Engineering (DFCE) in evaluation of the development and implementation of the Operational Civil Engineering-Air Force and Field Engineering Readiness Laboratory (OPS/CEAF FERL) as well as overall curriculum changes.

DFCE is implementing a non-traditional learning concept called "construct first, design later" developed by Col David Swint, DFCE Department Head which is quite different from the traditional model in which students are expected to understand general theoretical principles before working with specific applications. "Construct first, design later" asserts that doing should precede the study of theory. Thus, students are more likely to understand theoretical principles after hands-on experience with the application of those principles. If they repair a damaged runway, lay irrigation pipe, or place and finish a concrete pad, they will be better prepared, as a result of that experience, to understand the principles of theory-driven design.

The research involves a state-of-the-art comprehensive, multi-faceted evaluation approach which includes both formative and summative evaluation. The formative evaluation focuses on the process by which the program was developed and administered. The purpose of formative evaluation is to collect and use evaluation information as feedback for continuous improvement of the curriculum. The summative evaluation focuses on the outcomes of the program. The purpose of summative evaluation is to determine the changes in cadets' knowledge, skills and attitudes under the revised curriculum, and whether the objectives of the program were met. Note that the formative and summative evaluation approaches were conducted in parallel, and that both were intended to provide feedback to the CE faculty.

The procedures developed evaluate the impact of Civil Engineering 351 and are woven through the entire development and delivery of the course. Phase I of this newly developed five week summer course is conducted at various Air Force Bases in which cadets are exposed to Air Force civil engineering at the base level as well as exercises such as Red Horse and Silver Flag. Phase II of the course involves activities in the FERL in which there is a combination of academic and hands-on experiences. Phase III involves the integration of the OPS/CEAF and FERL

experiences into the CE curriculum. Thus, the entire CE curriculum is being redesigned to incorporate both the philosophy of "construct first, design later" and the actual experiences of cadets in the new summer program. The researchers collect qualitative and quantitative data by diaries, interviews, survey, and observational techniques while accompanying faculty and cadets to Eglin, Hurlburt, and Tyndall AFBs as well as Jack's Valley at the Academy. All instruments and procedures developed will eventually be transitioned to DFCE for their independent use.

Development of an Electronic Cockpit Map Display for Aircraft Ground Navigation

Anthony D. Andre, Ph.D.
Western Aerospace Labs
NASA Ames Research Center

David S. Tu
San Jose State University Foundation
NASA Ames Research Center

Abstract

The present study is part of a larger series of studies aimed at determining the potential of electronic cockpit taxi maps for improving the throughput and safety of low-visibility taxi operations. This study examined the relative benefits of: a) 3D (perspective) versus 2D (planar) map views, b) quantitative vs. qualitative heading information, c) constant vs. scaled aircraft icon sizes, and d) graphical route guidance. Twelve licensed pilots navigated 12 ground taxi routes using an electronic map display in the context of a B737 aircraft taxiing at San Francisco airport (SFO). The preliminary results show modest performance benefits for the 2D and 3D track-up maps over the 2D Overview, north-up map. Pilots preferred the 3D track-up map, the qualitative heading display, the scaled aircraft icon size and the graphical route guidance.

Low-visibility conditions present a host of problems for commercial aviation operations, especially when navigating on the airport surface. Under these conditions, the pilot's forward view is severely restricted, making it difficult to determine where the aircraft is, and where it should be going. Ironically, while many modern commercial aircraft are equipped to land (automatically) under low-visibility conditions, there is no such technology to aid in taxiing the aircraft from runway to gate, or vice versa (Andre, 1995). Consequently, flight throughput and sequencing is severely constrained, especially at the major airports. Recent efforts within NASA, the FAA, and the commercial aviation industry have been aimed at developing technologies to increase taxi speed and safety under low visibility conditions. The focus here is on the development of electronic cockpit taxi map displays to serve this purpose.

Andre (1995) recently conducted a study of pilot information requirements for low-visibility taxi operations while serving as a cockpit observer aboard thirty-five commercial carrier flights. Based on his cockpit observations, pilot interviews, and pilot-controller communication analyses, he concluded that an electronic cockpit taxi map could be an effective method for improving the throughput and safety of low-visibility taxi operations. Indeed, recent advances in display technology, global positioning (e.g., DGPS), and Datalink could combine to produce a taxi navigation display with potential benefits that far exceed those of stand alone paper charts (Batson and Hunt, 1994). However, for these displays to provide invariable assistance to the pilot, a careful, pilot-centered approach to their design and implementation must be undertaken (Andre, 1995). To this end, the criterion for deciding which map design features to implement should lie in whether or not these features increase a pilot's ability to maintain his/her awareness of where they are located on the airport surface (position), where they should be going (route), and where

other nearby aircraft are located (conflicts); the collective knowledge of which we refer to as "navigation awareness" (Aretz, 1991; Andre et al., 1991).

The present study examined 4 electronic taxi map design features in the context of a ground taxi simulation of San Francisco Airport (SFO). These features are: 1) Map Perspective (2D vs. 3D vs. Overview), 2) Heading Information (quantitative vs. qualitative), 3) Icon Size (scaled vs. constant) and 4) Graphical Route Guidance (on vs. off).

Method

Apparatus

The hardware used for the experiment consisted of a Silicon Graphics Indigo 2 Extreme workstation with two 21-inch color monitors, a BG systems flybox containing a 2-axis joystick for directional control, toggle levers for speed control and map zoom control, and a joystick trigger for selection of the map overview. The upper monitor showed the out-the-window view of SFO; the lower monitor showed the electronic map display.

Simulation

The upper screen consisted of a view of the San Francisco International airport surface environment. This view was similar to what the pilot would see out the cockpit window of a B737 jet aircraft. The view was complete with all relevant runways, taxiways, signs, gate markers and other landmarks. The taxi map display (see Fig. 1) was shown on the lower screen.

Map Perspective. An aircraft pilot, when looking out the forward window, navigates the aircraft from a self, or ego reference. Accordingly, it can be argued that a 3-D perspective taxi display would provide a more natural, or ecological representation of the airport environment than a conventional planar (2-D) display (Andre et al. 1991). Such a display would closely mimic the feedback provided the pilot from the forward visual scene, where the environment closer to the aircraft is represented more precisely than the environment farther away (Lasswell and Wickens, 1995). In the present study, a 3-D track-up taxi map was compared to a 2-D rendering of the same map. These two perspectives were compared to a 2-D Fixed (North-Up) map that could not be scaled (zoomed). This latter condition was meant to simulate the typical airport paper chart, which is the only current display aid available to pilots when taxiing (Andre, 1995).

Heading Display. Two types of heading displays were compared. The quantitative heading display consisted of a digital heading indicator located at the top of the map display. The combined heading display presented both a quantitative (digital) and a qualitative heading display; the latter consisted of four colored bars, each representing one of the cardinal directions (N,S,E,W). The bars rotated with the map display as the heading of the aircraft changed, thus directly indicating the direction the pilot would given any change in heading of the aircraft. This concept which showed success in a previous aircraft flight navigation study by Andre et al. (1991). It was expected that the combined heading display would allow pilots to make directional decisions faster and more accurately.

Icon Size. The aircraft icon was either scaled proportional to the map field of view (zoom level) or kept at a constant display size regardless of the field of view. On the one hand, the benefit of the scaled icon is that the true size of the aircraft relative to the map field of view is maintained, however at the larger map field-of-view settings, the icon may be so small that it is difficult to see. On the hand, the benefit of the constant display size is that at the larger map field-of-view settings (map zoomed out) the aircraft icon is maintained at a size that is still visible, although it may obscure (overlap) nearby information on the map.

Route Guidance. Route guidance was provided on 1/2 of the trials through the addition of a magenta line, roughly the width of a runway, connecting the starting point to the destination via the desired route. Based on a previous study using a similar simulation (Mejdal and Andre, in press), it was expected that pilots would complete the routes faster and more accurately when provided with graphical route guidance.

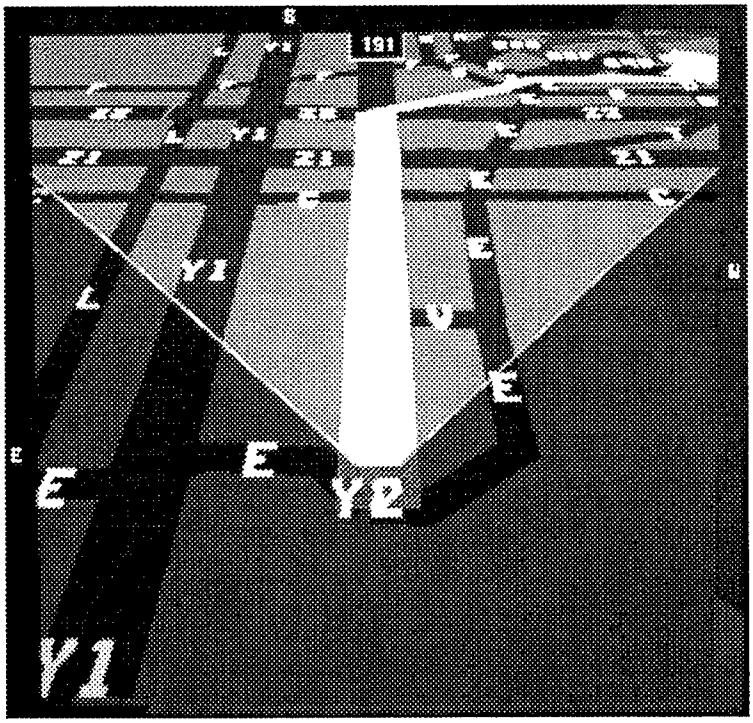


Figure 1. 3D electronic taxi map used in the study.

Wedge Display. Following the work of Aretz (1991) and Mejdal & Andre (in press), a wedge was always present on the map display. The wedge depicted the pilot's forward field of view on the map display.

Taxi Clearance Window. The taxi clearance instructions were located at the bottom left corner of the upper screen. Text directions to the desired runway or gate were provided in a format similar to the vocal clearances familiar to the pilots.

Controls. The speed of the aircraft was controlled by moving a toggle lever fore and aft—the further forward, the faster the aircraft traveled. A second toggle lever was used to control the zoom level of the map display. There were 6 different zoom (map field of view) levels. Pulling the lever back gave a higher level view (zoomed out), while moving the lever forward gave a lower level view (zoomed in). The joystick controlled the front wheels of the aircraft thus enabling the pilot to turn. The joystick also featured a "trigger" that, when depressed, was used bring up a north-up, overview map of the airport surface.

Procedure

During all trials the visibility was set to approx. 700 ft. RVR resulting in a view that was severely limited due to simulated fog. All factors were varied within subjects.

Post-Test Survey

At the end of the series of 24 trials, the subjects were given a survey questionnaire. Questions concerning their preference between map displays and the navigational aids were given. They were asked to describe when and why each map feature was helpful to them.

Subjects

Twelve licensed general aviation pilots were paid to participate in the study.

Results

Pilot Performance

The data show reduced (faster) route completion times for the 2D and 3D track-up maps relative to the 2D north-up overview map. In addition, pilots made fewer navigational errors when using the 2D and 3D track-up maps compared to the 2D north-up overview map. Pilots planned and completed the routes faster with graphical route guidance and made fewer navigational errors relative to the unguided condition. Further, there appears to be no cost to the pilots' ability to respond to unexpected events (e.g., incursions) when using the route guidance, in contrast to previously findings (e.g., Mejdal and Andre, in press). There was no effect of the icon size or heading information manipulations.

Pilot Preferences

Sixty-seven percent (8/12) of the pilots preferred the 3D map the most. Twenty-five percent preferred the 2D map the most and only one pilot (8%) preferred the overview map the most. Eighty-three percent (10/12) of the pilots toggled on the overview map at some time while using the 2D or 3D maps. Seventy-five percent (9/12) of the pilots preferred the qualitative heading display over the quantitative (digital only) heading display. Eighty-two percent (9/11) of the pilots preferred the scaled aircraft icon size over the constant aircraft icon size. Sixty-seven percent (8/12) of the pilots stated that the wedge display (Aretz, 1991) was useful. Finally, all the pilots (12/12) stated that the route guidance information was useful.

Discussion/Conclusion

This study examined four taxi map display features for improving the capability, safety and efficiency of low-visibility taxi operations. The results demonstrated the general benefit of the moving map display during low-visibility conditions, as the pilots were able to taxi at speeds and accuracy levels similar to that of high-visibility conditions. Consistent with previous results (Mejdal and Andre, in press), the results showed a performance benefit for the track-up maps and the addition of graphical route guidance. Further, the majority of pilots preferred the 3D map, the qualitative heading display, the scaled aircraft icon size and the graphical route guidance. Based on these preliminary results, the 3D (perspective) taxi map appears to be a viable display option and warrants further investigation.

References

Andre, A.D. (1995, in press). Information requirements for low-visibility taxi operations: What pilots say. To appear in Proceedings of the 8th International Symposium on Aviation Psychology. Columbus, OH: The Ohio State University.

Andre, A.D., Wickens, C.D., Moorman, L., & Boschelli, M.M. (1991). Display formatting techniques for improving situation awareness in the aircraft cockpit. The International Journal of Aviation Psychology, 1(3), 205-18.

Aretz, A.J. (1991). The design of electronic map displays. Human Factors, 33, 85-101.

Batson, V.M., Harris, R.L., & Hunt, P.J. (1994). Navigating the Airport Surface: Electronic vs. Paper Maps. Presented at the 13th Digital Avionics Systems Conference, Airport Surface Operations. Phoenix, AZ.

Lasswell, J. W. & Wickens, C.D. (1995). The effects of display location and dimensionality on taxi-way navigation. Aviation Research Laboratory Technical Report ARL-95-5/NASA-95-2. Savoy, IL: University of Illinois.

Mejdal, S. & Andre, A.D. (1996, in press). An evaluation of electronic map display features for aircraft ground navigation. To appear in proceedings of the 1996 ErgoCon Conference. San Jose, CA: Silicon Valley Ergonomics Institute.

Airsickness During Flight Training

Thomas G. Dobie MD PhD
University of New Orleans, University of Leeds (UK)
and Naval Biodynamics Laboratory, New Orleans
James G. May MS PhD
University of New Orleans

Abstract

This paper concerns the problem of airsickness during flight training. The reader will be given an overview of motion sickness; its incidence, etiology, and management. This includes the physiological mechanisms underlying motion sickness and the psychological mechanisms which aggravate this condition. It also discusses the form of therapy known as cognitive-behavioral training, based on a technique first described by Dobie (1965). This technique focuses on the psychological aspects of stress management and tries to encourage a sense of confidence in the individual so that he or she can tolerate noxious or stressful situations. This belief, once established, is reinforced with controlled exposures to provocative motion. Although the technique appears to involve habituation and adaptation to a particular situation, mere repetitive exposure to a provocative motion environment, without counseling, has not proven to be beneficial. While this technique appears to include many elements often seen in the management of neurotic disorders, the procedures were not developed within the framework of a mental health model. The emphasis is always on the normality of this protective response to provocative motion situations.

Airsickness is common during flight training and can be dealt with successfully. Many would suggest that the solution lies in the field of selection so as to avoid the problem. Our experience is that selection is not the simple solution that it might appear. Perhaps more important is the fact that such a procedure might well exclude volunteers who would otherwise be above average students and excellent operational flight crews.

Definition of Motion Sickness

Motion sickness is a response to real or apparent motion to which a person is not adapted. It is characterized by malaise, general discomfort, pallor, sweating, nausea and vomiting. Provocative motion environments involve many forms of transport, such as aircraft, ships, and automobiles. Motion sickness is also experienced in flight simulators and the microgravity of space shuttle missions.

The term motion sickness is, however, a complete misnomer for this response. "Sickness" or "illness" suggests that there is something wrong with the individual; that they are suffering from some kind of malady. In truth, that person is exhibiting a number of physical signs and symptoms

of a bodily disturbance, but these are the result of a built-in protective response caused by exposure to provocative motion environments for a sufficient length of time. It is only the stimulus, or environment, which is abnormal and not the person.

Symptoms and Signs of Motion Sickness

The main symptom of motion sickness is nausea and the main signs are pallor, sweating and vomiting. However there are many other responses: such as, apathy, general discomfort, headache, stomach awareness, increased salivation and prostration. More important than particular symptoms, however, is the deleterious effect that motion sickness has on performance.

Incidence of Motion Sickness

The incidence of motion sickness is extremely variable depending upon the circumstances. Rubin (1942) quoted a figure of 11% (ranging from 6% to 22% with different training courses), for the incidence of airsickness during basic flight training. A survey of flight instructors' post-flight reports of 577 RAF flight trainees showed that 38.7% suffered from airsickness at some time during their basic flight training on single-engine jet aircraft, usually in the early stages (Dobie, 1974). In more than a third of these cases it was severe and protracted and had a detrimental effect on training effectiveness. A study of US Navy officers undergoing flight training for various non-pilot crew duties revealed a mean incidence of airsickness in 13.5% of all flights. This was judged to have caused a decrement in the trainees' performance in 7.3% of flights (Hixson, Guedry and Lentz, 1984).

Etiology of Motion Sickness

Although the mechanism has not yet been determined with absolute certainty, changing acceleration acting on the labyrinth is clearly a basic cause of motion sickness. However, "motion sickness" can be caused by purely visual stimulation, without associated bodily accelerations, as well as by motion which causes changing linear and angular accelerations. Adaptation to motion also occurs (given sufficient time) and a motion sickness response can then occur when the adapted person returns to the normal motion environment. All these features must be taken into account when explaining the underlying cause or causes of motion sickness.

Physiological Motion Underlying Motion Sickness

The currently most acceptable explanation of motion sickness is that the physiological component is the body's response to inharmonious sensory information reaching the so-called "comparator" in the brain. The motion stimuli originating from active or passive bodily motion are mainly detected by the eyes and the vestibular apparatus. Additionally, however, changes in the body's orientation to the gravitational field and other added linear accelerations can also stimulate mechanoreceptors in the body located in the skin, muscles, joints and other tissues.

This physiological explanation for motion sickness is called "sensory conflict", indicating that there is some sustained disynchrony at the level of the "comparator" in the brain, (Reason, 1978; Oman, 1982). Not only might the incoming signals be in conflict with each other, they might also be in disagreement with those which the brain expects to receive.

Psychological Mechanisms Which Exacerbate Motion Sickness

There is also a psychological component to the causation of motion sickness. It is natural to develop an anxiety due to feelings of discomfort or nausea brought about by certain provocative maneuvers, or when exposed to a different and unfamiliar mode of travel. This is due to the arousal which typically develops when one is exposed to situations which are known to be uncomfortable or threatening. Personality differences might also determine how an individual reacts to these motion discomforts, in terms of anticipation and/or severity of response. This does not in any way imply that motion sickness is a neurotic response on the part of that individual. On the contrary, this is seen as a perfectly normal and understandable "protective" response. Indeed RAF flight trainees who were successfully treated for apparently intractable motion sickness appeared to be high achievers who performed particularly well on their return to full flight status (Dobie, 1974).

In summary, the underlying cause of motion sickness is likely to be due to a form of sensory mismatch, together with an individual's experiential anxiety caused by that individual's attitudes, memories and past experiences with motion stimuli.

Treatment of Motion Sickness

We have stressed that the term motion sickness is misleading but it continues to be used because, regrettably, it has become the accepted term. This is not just a question of semantics, however, because the terms "motion sickness" or "motion illness" by their very nature may well account for the fact that the main approach to the treatment of this response has been pharmacological, in the classical mode of dealing with an "illness". However, this is not necessarily the best approach in many circumstances.

The Pharmacological Approach to Treatment

This brief review sets out to explore briefly the shortcomings of this approach particularly in terms of the skilled operator (rather than the passenger). The pharmacological approach to the treatment of motion sickness introduces many problems. The drug actions are variable both in terms of individual responses and the effects of the operational situation on these responses. Some of the potential side effects are not acceptable when the individual is in control of sophisticated equipment or complex operational command and control situations. For example, flying is both a skilled and potentially dangerous occupation, so that any decrement of performance brought about by medication can be very serious. The use of an anti-motion sickness drug should be restricted to those situations where the trainee is flying dual and therefore not in sole charge of the aircraft, nor responsible for a critical task in the air. Nor should physicians prescribe anti-motion sickness medication to flight crews for long periods, lest the user

becomes dependent upon it. Many individuals who have grown used to the protection of an anti-motion sickness drug are known to be apprehensive about flying without appropriate medication. In summary, the pharmacological approach is neither simple nor straight-forward.

The Use of Non-Pharmacological Therapy

A number of different forms of therapy have been developed in various centers around the world for the treatment of motion sickness without recourse to medications. Some of these different approaches to desensitization use a variety of devices, others include biofeedback, whereas cognitive-behavioral training relies on one piece of equipment only, with supporting counseling sessions.

Dobie first instituted this program of cognitive-behavioral therapy during the period 1960 - 1970 in order to help flight trainees in RAF Flying Training Command who were suffering from airsickness. A considerable amount of flight training time was being lost and in the very worst cases of motion sickness, flight trainees were being permanently grounded. Many questions arose concerning the prevention and treatment of airsickness, but at that time not many hard answers were available other than the restricted use of medications.

The majority of people who suffer from airsickness when they first start flying adapt to the new environment within the first 15 hours or so of flying and their symptoms disappear. This time scale varies with the breakdown of the flight training program and type of aircraft. Some student aviators have a more prolonged history of airsickness and need further help and encouragement. A smaller but very important group of trainees fails to respond to early treatment despite the efforts of the flight instructors and medical officer and reach the stage of becoming intractably airsick. The decrement of performance in these students can be so severe that it critically affects their progress and their training supervisors must decide whether or not it is justifiable to allow them to continue flight training.

Intractable airsickness represents a large economic loss to the flight training organization. Not only are these highly motivated and potentially valuable people on the verge of becoming training failures, they have already cost a large amount of money in terms of training hours and supervisors' time. For example, Jones et al. (1985) estimated the loss of a student pilot at 15 hours to be over $15,000 and a trained flier around the half-million dollar mark. The figures are much higher today. In 1994, a figure of 3 million pounds sterling has been suggested as the value of an experienced front-line RAF pilot. So, from the point of view of cost-effectiveness, a successful anti-motion sickness program has great merit.

The Early Cognitive-Behavioral Approach - RAF 1960-1972

A person suffering from severe incapacitating motion sickness inevitably shows some degree of anxiety or loss of confidence by the time he or she is referred for a second opinion. This psychological overlay is inevitable, due to anticipatory anxiety associated with provocative motion stimuli which have previously caused motion sickness. In addition, "professionals" who experience motion sickness feel that their future career is in jeopardy, which adds to the "arousal".

Dobie, therefore, decided to base his form of treatment, now known as cognitive-behavioral therapy, on vestibular training as a means of desensitization, together with confidence-building counseling (Dobie, 1974).

The desensitization element consists of building acclimatization to vestibular stimulation on a rotating/tilting chair. The passive head movements involved produce cross-coupled or Coriolis stimulation of the semicircular canals resulting in a sensation which is frequently bizarre and disorienting. The stimuli are carefully controlled so that subjects never experience more than the early symptoms of motion sickness and no one ever gets even close to emesis. This moderate approach is critical to the development of confidence. The technique addresses the main problems in parallel, namely lack of acclimatization to motion and a heightened arousal. A candidate's improved performance on the rotating/tilting table, shown by an ability to withstand increasing amounts of vestibular stimulation over time, helps to increase confidence and lessen anxiety.

Table I

RESULTS OF COGNITIVE-BEHAVIORAL TRAININ IN THE RAF PRIOR TO 1974

CLASS	TOTAL	PASS	FAIL	
			NOT AIRSICK	AIRSICK
STUDENT AIRCREW	44	34	4	6
QUALIFIED AIRCREW	6	4	1	1
ALL	50	38	5	7
% TOTALS	*100*		*86* (SUCCESS)	*14* (FAILURE)

Results of Treatment

The overall results of this program showed that all individuals made improvements in their tolerance to stimulation on the motion device and 86% of them were successfully returned to full unrestricted flying. Subsequently, five of these subjects (10%) did fail flight training, but they did so for reasons which their flight instructors and executive supervisors confirmed to be totally unrelated to motion sickness (Table I).

It should be stressed that a 10% failure rate at that stage of training was significantly lower than usual. This seemed to indicate that the trainees who were treated for intractable airsickness were above average students. That conclusion was supported by a long-term follow-up, some six years later, by which time the candidates were on operational squadrons. The follow-up confirmed the successful retention of all of our subjects who had completed training. In addition, this group of individuals was rated above the average. It also confirmed that they were no longer hampered by motion sickness.

Evaluation of Key Compnents of Cognitive-Behavioral Training

When Dobie first reported this procedure, queries were raised concerning the need for the cognitive component. It was suggested that the effectiveness of the program was perhaps due to behavioral desensitization alone. Recently, we investigated the relative importance of these two factors: the counseling component and behavioral desensitization (Dobie et. al., 1989). The results indicated that only the two groups receiving cognitive counseling (counseling only) and (counseling plus desensitization) demonstrated significant pre- to post-test increases in tolerance to visually-induced apparent motion and decreases in motion sickness symptomatology.

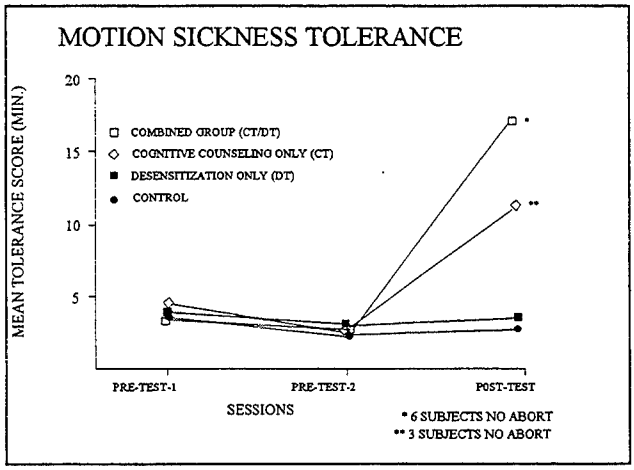


Figure 1.

Furthermore, the group receiving both counseling and desensitization (cognitive-behavioral training) showed significantly greater tolerance than counseling only group. The group receiving desensitization only did not differ from control (Figure 1). These results indicate that mere repeated exposure to the provocative stimulation is not sufficient to reduce motion sickness. This emphasizes the importance of cognitive factors in motion sickness.

Review of Military Desensitization Programs

It is difficult to compare the results of the different military desensitization programs because, unlike Dobie's original program, each of the others includes some form of pre-selection. In addition, the other programs are more complex. For example, the USAF biofeedback program and the Canadian Forces (CF) airsickness rehabilitation program require additional instrumentation to record electromyographic data for biofeedback training. The current RAF program also uses linear Gz oscillation and angular oscillation in addition to the cross-coupled stimulation used in the cognitive-behavioral training program. In terms of rehabilitation flight time, the CF program normally consists of six flights in a basic jet training aircraft, which is similar to the early RAF proposals. In the RAF this has been increased to include special rehabilitation in a designated aircraft.

The published results obtained during the three phases of the RAF program, namely Dobie's original pre-1974 program, (Dobie, 1974), the interim years 1974-1980 and finally Bagshaw and Stott's 1981-1983 program, (Bagshaw and Stott, 1985), are shown in Table II, along with those from the USAF and the CF (Banks, Salisbury and Ceresia, 1992).

Table II

SUMMARY OF MILITARY DESENSITIZATION PROGRAMS, BASE ON BANKS et al., WITH THE ADDITION OF THE ORIGINAL COGNITIVE-BEHAVIORAL RESULTS IN THE RAF PRIOR TO 1974

PROGRAM YEARS	RAF (PRE 74)	RAF (74-80)	RAF (81-83)	USAF (79-85)	CF (81-91)
TOTALS	N = 50	N = 46	N = 32	N = 34	N = 22
% SUCCESSFULLY DESENSITIZED	76	67	72	62	54.5
*% SUCCESSFULLY DESENSITIZED	10	02	12.5	14.7	22.7
% TOTAL SUCCESS	86	70	84	76.5	77.3
% TOTAL FAILURE	14	30	16	23.5	22.1

* FAILED FLIGHT TRAINING FOR REASONS OTHER THAN MOTION SICKNESS

It is evident that all of the programs are effective. However, none of these newer programs has apparently improved upon the success rate of the original program, despite the additional efforts and extra costs involved. This calls into question the value of complicating the relative simplicity of the original cognitive-behavioral approach, quite apart from the significant increase in the cost involved in so doing.

Conclusion

Motion sickness is a very common and debilitating response to provocative motion environments. It is a normal protective mechanism and not a neurotic response. Apart from any physiological differences between individuals, which are difficult to detect, it involves for the most part a cognitive overlay based on previous motion experiences.

Anti-motion sickness drugs which are effective in reducing or preventing symptoms generally exhibit undesirable side effects. For that reason they are not suitable for situations in which the motion susceptible individual is required to perform skilled tasks or is in control of potentially dangerous equipment. Cognitive-behavioral training, on the other hand, is also effective but carries no such penalty in terms of side effects.

Cognitive-behavioral training focuses on the psychological aspects of stress management. The technique endeavors to instill a belief that the individual can tolerate noxious or stressful situations. Once established, this belief is reinforced with controlled exposures to provocative motion stimulation. While the technique appears to involve habituation and adaptation to a particular situation, mere repetitive exposure without counseling has not proven to be beneficial. A key element in the technique appears to be the individual's ability to learn to control cognitive focus. The counseling procedures have been applied successfully to various kinds of motion sickness.

Cognitive-behavioral anti-motion sickness training has not only been shown to be very successful, more important still, it has saved the careers of highly motivated successful flight crews who would otherwise have been lost to the military. Airsickness is an important problem which affects flight training organisations world-wide. It should be recognised for what it is,

namely, a normal protective response to an abnormal motion environment - and it can be treated successfully - saving some of the best trainees available.

References

Bagshaw, M, Stott, JRR. The Desensitisation of Chronically Motion Sick Aircrew in the Royal Air Force. Aviat. Space Environ. Med. 1985; 56, 1144-1151.

Banks, RD, Salisbury, DA, Ceresia, PJ. The Canadian Forces Airsickness Rehabilitation Program. Aviat. Space Environ. Med. 1992; 63: 1098-101.

Dobie, TG. Motion Sickness during Flying Training. AGARD Conf. Proc. 1965; 2: 23-32.

Dobie, TG. Airsickness in Aircrew. AGARD/NATO, AG-177, Neuilly-sur-Seine, France. 1974.

Dobie, TG, May, JG, Fisher, WD, Bologna, NB. An Evaluation of Cognitive-Behavioral Therapy for Training Resistance to Visually-Induced Motion Sickness. Aviat. Space Environ. Med. 1989; 60: 307-14.

Hixson, WC, Guedry, FE, Lentz, JM. Results of a Longitudinal Study of Airsickness Incidence During Naval Flight Officer Training. In: Motion Sickness: Mechanisms, Prediction, Prevention and Treatment. Conf. Proc., 372. AGARD/NATO, Neuilly-sur-Seine, France. 1984.

Jones, DR, Levy, RA, Gardner, L. Marsh, RW, Patterson, JC. Self-Control of Psychophysiologic Response to Motion Stress: Using Biofeedback to Treat Airsickness. Aviat. Space Environ. Med. 1985; 56:
1152-7.

Oman CM. A Heuristic Mathematical Model for the Dynamics of Sensory Conflict and Motion Sickness. Acta oto-laryngol, suppl. 392. 1982.

Reason, JT. Motion Sickness Adaptation: A Neural Mismatch Model. J. R. Soc. Med. 1978; 71: 819-29.

Rubin, HJ. Airsickness in a Primary Air Force Training Detachment. J. Aviation Med. 1942; 13: 272-6.

An Operational Definition and Measurement Method for Situation Awareness

Bruce P. Hunn,
US Air Force Flight Test Center,
Edwards Air Force Base, California.

Situation awareness has been defined from a theoretical perspective for a number of years and is considered as both a process and a product, however, there have been few operationally testable, definitions of Situation Awareness proposed. Historically, several techniques have been employed to measure situation awareness but those techniques were borrowed from the measurement of mental workload, which is associated with, but not synonymous to situation awareness. The purpose of this discussion is to define situation awareness in a way which is operationally significant and discuss ways in which it can be operationally tested.

Introduction

Situation Awareness has been defined repeatedly (Endsley, 1987, 1990, 1995), and current definitions contain perception, comprehension, and projection as elements. Perception is primarily concerned with the awareness half of SA. The human-in-the-loop must be aware and able to perceive informational cues in order to process those cues into the next stage of comprehension. In perception, "status, attributes and dynamics of relevant elements in the environment" are observed and held in Short Term Memory (Endsley, 1995). In the comprehension stage diverse data from short term memory is converted into significant information for the user of the system.

In application, SA has been measured in a variety of circumstances beginning with the assessment tool SAGAT, "Situation Awareness Global Assessment Technique", which was developed using an aircraft simulator as an assessment tool (Endsley, 1987). Essentially, a simulator involved the pilot in a detailed flight scenario which was abruptly stopped and the pilot was then questioned regarding the flight situation presented. Pilot recall was compared to actual simulator status records and the degree to which the pilot could recall flight parameter detail was represented as a measure of the pilot's SA. While providing information on pilot recall ability, the use of SAGAT did not directly answer the question of the relative importance of particular information to the pilot, and how that information was critical to the pilots task performance and level of SA. In addition, the process had very limited practical application outside a simulator, since there is no way to put time on hold for assessment purposes. SAGAT did provide a good measure of global information awareness which is a critical step in determining what information was readily accessible to the pilot.

Measures originally designed for mental workload assessment like, P300 EEG, external task performance, embedded task performance, self ratings, subjective ratings, and external observer ratings have been used on SA. The problem with most of these measures is that while they provided some utility in measuring mental workload, SA is not synonymous with mental workload. While there are similarities and interactions between the two constructs they are not interchangeable in terms of assessment, in fact, some of the SA measurement techniques, such as observer ratings, may even have very poor face validity. The question of how an external

observer can devine the dynamic mental state, and level of situation awareness of another person is problematic to say the least.

Situation Awareness Theory

Several of the conclusions of Smith and Hancock, (1995) are particularly appropriate to a discussion of SA. What Smith and Hancock propose is that the human-in-the-loop must develop a "level of adaptive capability sufficient to match the specification of task goals". It is this view of SA which may be key in determining how to operationally define and measure it. They also propose that "only with a specified goal and concrete performance criteria can we begin to talk about how well adapted a particular agent is with respect to a particular environment."(Smith and Hancock, 1995).

Smith and Hancock also state that SA is the competence which controls behavior, and further consider it an invariant of the individual. Their use of the term "invariant" relies on the perceptual model proposed by Neisser in 1976. Neisser's model shows the human to be the focal point in a process which involves sampling information, creating a cognitive model and then directing further information sampling to arrive at a refined and therefor more appropriate model. For purposes of this discussion, the invariant part of the individual could also be compared to that individual's schemata, or mental models which will, in turn, direct their response behavior. An excellent discussion of these items may be found in Frederico, (1995), and Adams, Tenney, and Pew, (1995). Rather than discussing the structure of mental models and cognitive sampling the focus of this paper will be on assessing behaviorally quantifiable aspects of SA.

A second item which relates to SA is the idea of "risk space" (Smith and Hancock, 1995) or "decision event" as discussed by Orasanu and Connolly, (1993). This concept indicates that for the human-in-the-loop there are certain self imposed boundaries which they establish a-priori to the performance of a task, for example these boundaries could be the depth to which they seek information or the amount of time which they spend in seeking information. An assessment of risk space could easily lend itself to a quantitative behavioral measurement approach.

It is also important to not directly relate the quantity of a particular type of behavior (like information sampling rates) with the quality of the final system performance. To associate individual behaviors directly to an SA measure would lead to errors in attribution, described by Flach (1995) where SA would be seen not as an intervening variable but as a direct causal agent. Information sampling and SA is probably a correlational, rather than directly causal. Accident investigation often implies that the event occurred due to a loss in SA, but the accident's occurrence could have been dependent on a faulty mental model or inadequate schemata which misorganized critical information. In most accidents, a lack of attention to a particular item is usually not the only factor, it is one of a combination of factors. This combination of factors introduces the third segment of SA assessment, prioritization.

Every pilot realizes that prioritization of decisions and action is the key to successful flight. Time and SA are closely tied to each other since a situation is by definition, a transient event in time. In addition, prioritization is dependent on the knowledge level of the operator,

since without system knowledge no adequate prioritization scheme can be created. The temporal aspects of SA are discussed by Sarter and Woods (1991) and the implications of this temporal sampling on the subjects invariant or "mental picture" is further refined by Dominguez (1994). The second feature of this temporal discussion is that the best choice may be "an adequate alternative" rather than an optimal one (Federico, 1995, Beach and Lipshitz, 1993, Orasanu and Connolly, 1993). This prioritization, or strategic sampling over time, is a key to determining what SA is and secondly, how to measure it. A succinct review of strategic sampling is found in Salas, Prince, Baker and Shrestha (1995).

Based on the previous discussion of an individuals "invariant" behavior, and definitions of risk space and prioritization, an acceptable operational definition of SA might be: "An individual's selection of the appropriate level of risk for a task and the individuals prioritization of information for the completion of that task". While this definition of SA is based on cognitive structures (i.e. invariants, schemata, mental models) it is also based on assessing SA through observable behavioral measures

Situation Awareness Testing Methodology

The first step in the measurement of SA is in an a priori definition of the goals of the tasks being set by the test administrator, i.e., task completion time, number of errors allowed, subtask completion importance, etc. This external (test administrator) risk space definition will be followed by the subject detailing their own risk space and prioritization criteria for the task. The task should be performed by subjects who have performed a similar task before (and therefor have system knowledge) but should not be so routine as to become automatic in nature. Each subject will then create their own "profile" or estimation of what the task will involve before attempting it. This will include two important items; 1. The "invariant" or characteristic approach to the problem by the subject. 2. Their individual assessment of the risk associated with performance of the task.

One of the assessment tools used for the test could involve the subjects providing time estimates for the completion of tasks, estimates of expected error rates, etc. The subjects data from this test can then be compared to other subjects and a correlational pattern may be observed. These correlation's should reveal what factors are driving Situation Awareness. In other words, rather than making a post-test assessment of what constitutes SA, an a-priori estimate should be made as well.

What is inferred with this method is that an a priori estimate of performance requires an awareness of the elements of a situation. In other words, estimates of future occurrences cannot be accurately made unless there is an understanding of the elements which may constitute that event. In addition, all events are represented in the framework of a person's mental models, but those mental models can only be inferred and not directly observed. What can be observed are the behavioral actions associated with task performance under controlled test conditions. What is critical to this discussion is not how accurate the estimates or performance are, but the structure of how the subject is searching for situational information.

Conclusions

The previous sections have briefly reviewed Situation Awareness and outlined a definition of SA based on behaviorally measurable criteria. This method of assessment of SA involved; determining specific task goals, understanding the subjects "invariant" response pattern, measuring risk space parameters and evaluating prioritization of task critical information, both before and during task performance. The SA definition proposed above also used simple, behavioral, and quantifiable data for the determination of SA in operational settings. While use of this definition relies on theoretical concepts of "mental models", "invariants" "decision events" and other cognitive constructs it ultimately uses common, observable, behavioral data for its proof. The focus of this definition and methodology is to address SA as an intervening variable which has a correlational (rather than causal) relationship to human performance.

The proposed method used for assessment of SA also relies on the correlational relationships of, risk and prioritization being measured before, and during the performance of an adaptive task. The assessment of the subject's level of SA a-priori to beginning the task and the correlation of those measures with their task performance is one of the salient features of this method and distinguishes it from other SA assessment methods. Evaluation of a process oriented construct like SA both prior to and during its application should have a greater degree of face validity than using measures designed to assess SA via task performance after its completion. In the assessment of situation awareness, the goal is not to provide an assessment of a performance product as much as it is to determine the process used to achieve that product.

References

Adams, M.J., Tenney, Y.J., Pew, R.W., (1995). Situation Awareness and the Cognitive Management of Complex Systems. Human Factors, 37(1), p 85-104. The Journal of the Human Factors and Ergonomics Society. Santa Monica, Ca.

Beach, L. and Lipschitz, R., (1993). Why classical decision theory is an inappropriate standard for evaluating and aiding most human decision making. In G. Klein, J. Orasanu, R. Calderwood, and C. Zsambok (Eds.), Decision making in action: Models and methods (pp.21-35). Norwood, NJ: Ablex.

Dominguez, C. Can SA be defined? (1994). In M. Vidulich, C. Dominguez, E. Vogel, and G. Mcmillian (Eds.), Situation Awareness: Papers and annotated bibliography (pp.5-15); Report AL/CF-TR-1994-0085). Wright-Patterson Air Force Base, OH: Air Force Systems Command.

Endsley, M.R. (1987). SAGAT: A methodology for the measurement of situation awareness (NOR DOC 87-83). Hawthorne, CA: Northrop Corp.

Endsley, M.R. (1990). A methodology for the objective measurement of situation awareness. In situation awareness in aerospace operations (AGARD-CP-478; pp1/1-1/9) Neuilly-Sur-Seine, France: NATO- Advisory Group for Aerospace Research and Development.

Endsley, M.R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. Human Factors, 37(1), p 32-64. The Journal of the Human Factors and Ergonomics Society. Santa Monica, Ca.

Federico, P,A. (1995). Expert and Novice Recognition of Similar Situations, Human Factors, 37(1), p 105-122. The Journal of the Human Factors and Ergonomics Society. Santa Monica, Ca.

Flach, J.M. (1995). Situation Awareness: Proceed with Caution. Human Factors, 37(1), p 149-157. The Journal of the Human Factors and Ergonomics Society. Santa Monica, Ca.

Neisser, U. (1976). Cognition and reality: Principles and implications of cognitive psychology. San Francisco: Freeman.

Orasanu, J.M. and Connolly, T. (1993). The reinvention of decision making. In G. Klein, J. Orasanu, R. Calderwood, and C. Zsambok (Eds.), Decision making in action: Models and methods (pp.3-20). Norwood, NJ: Ablex.

Salas, E., Prince, C., Baker, D.P., and Shrestha, L. (1995). Situation Awareness in Team Performance: Implications for Measurement and Training Human Factors, 37(1), p 123-136. The Journal of the Human Factors and Ergonomics Society. Santa Monica, Ca.

Sarter, N.B., and Woods, D.D. (1991). Situation awareness: A critical but ill-defined phenomenon. International Journal of Aviation Psychology, 1, (pp.45-57).

Smith, K., and Hancock, P.A. (1995). Situation Awareness Is Adaptive, Externally Directed Consciousness. Human Factors, 37(1), p 137-148. The Journal of the Human Factors and Ergonomics Society. Santa Monica, Ca.

Subjective Workload Measures: National Aeronautics and Space Administration Task-Load Index in a Task-Saturated Cockpit Environment

Keith R. Ober
Anthony J. Aretz
United States Air Force Academy

Abstract

The National Aeronautics and Space Administration Task-Load Index (NASA TLX) was explored for its sensitivity in high workload situations. An Israeli Instrument Pilot Evaluation System (PES) was used to collect TLX ratings in varying high workload conditions with 27 Air Force Academy cadets as subjects. TLX ratings were regressed on task conditions to evaluate the measure in terms of a theory proposed by Yeh & Wickens (1988). In contrast to predictions, the data suggest the NASA TLX is sensitive to workload variation in high workload situations. In contrast to existing theory, frustration seemed to have a negligible affect on workload ratings, even on overloaded subjects. The number of concurrent tasks, the number of conflicts in each stage of information processing, and temporal demands were significant predictors of workload and simulator performance. Implications for future studies are discussed.

Subjective ratings of perceived workload have gained both popularity and experimental support due to their ease of use and face validity. Subjective workload ratings have been tested against other types of workload measures, against each other, and against themselves; all comparisons indicate the ratings are valid and reliable (e.g., Hill, Iavecchia, Byers, Bittner, Zaklad, and Christ, 1992). The growing body of knowledge concerning subjective ratings of perceived workload has produced several broad theories.

One theory proposed by Yeh and Wickens (1988) is based on the multiple resource theory developed by Wickens (1992). One of the key components of this theory was that subjective reports of workload will be most affected by changes in demands on working memory. The implication of this axiom is that subjective workload measures will be insensitive to changes in actual workload if working memory is already completely "tapped." In other words, if working memory is overloaded, subjective ratings of changes in workload will be solely a function of subject frustration, not actual workload changes. A final contribution of Yeh and Wickens (1988) was the important relationship between workload and performance: their association (or disassociation) depends on the efficiency with which multiple tasks are time-shared. According to the theory, if multiple tasks are efficiently time shared, performance can be predicted by subjective workload measures.

These predictions were supported by Aretz, Shacklett, Acquaro, and Miller (1995) who found that the number of concurrent tasks (up to five concurrent tasks) was a primary contributor to workload measures. In addition, they found that the "effort" dimension of the NASA Task-Load Index (TLX) subjective workload measure was the most significant aspect of workload

ratings. In contrast, Yeh and Wickens (1988) suggested that subjective measures would be more indicative of frustration levels in high workload conditions. A second aspect of the study that suggests further analysis is the relationship between multiple resource theory and subjective workload ratings. Yeh and Wickens suggest that multiple resource conflicts play an important role in subjective workload ratings; however, resource conflicts were not examined in Aretz et. al. study.

Though there is a large body of theory concerning subjective workload measures, there is a dearth of experimental data regarding the performance of subjective workload measures in high workload situations. The theory of Yeh and Wickens implies that workload measures will be ineffective when working memory is fully taxed; however, conclusions of Aretz et. al. (1995) indicate that subjective measures may in fact be sensitive to workload changes even in a task-saturated environment.

Another factor in the correlation between workload ratings and performance is the type of tasks concurrently performed. According to Wickens (1992), there are three stages in which information processing may cause performance decrements: input channel conflicts, central processing conflicts, and response channel conflicts. Input channels can be either visual or auditory, central processing can be either spatial or verbal, and responses can be either manual or verbal. For example, if two concurrently performed tasks both require spatial central processing (mental rotation and catching a baseball, for example), they will have a central processing conflict and performance should suffer; however, it is unclear whether these conflicts influence subjective workload ratings in task saturated conditions.

The present study was designed to assess the sensitivity of the NASA TLX in high workload situations from a multiple resource theory perspective, trying to provide additional support for the proposition of Yeh and Wickens (1988) that working memory demands are the greatest contributor to workload ratings. Analysis of different types of concurrent tasks will seek to clarify the role of different types of information processing conflicts on workload ratings. In addition, the present study examined the suggestion by Yeh and Wickens (1988) that frustration will be the main contributor to workload rating variance in high workload situations. The NASA TLX sub-scales will also be useful in diagnosing the important contributors to overall workload ratings.

Method

Subjects

Subjects in the present study included 27 Air Force Academy cadets (24 male), age 17 to 22. All but 9 of the subjects had less than 2 hours of powered aircraft flight time; of the 9 with flight time, the mean was 44.7 hours with a range of 2 to 106 hours.

Apparatus

Flight Simulator. The flight simulator was an Israeli Pilot Evaluation System (PES) developed by Israeli Aircraft Industries (IAI). The PES consists of a 486 computer, a partially enclosed "cockpit" (including a radar display screen, headphones, and F-16 type throttle and stick), and PES software (see Figure 1). In addition to the standard PES hardware, a tape player was also used to present an auditory digit subtraction task. A short computer generated slide show, provided by IAI, was used for training prior to PES evaluation. The flight scenarios were configured to present from 1 to 6 concurrent tasks, including the digit task. The sub-tasks are described below.

The Target Intercept was identified as the "primary task" to the subjects. A target appeared as a square on the radar screen. The subject had to move a cursor over the target and lock on the target. After locking on the target, the subject had to disengage and re-engage the target at an 18 mile range. When the target flew into the missile launch envelope, the subject had to arm a missile and shoot. If a missile was fired within the correct parameters, the target would

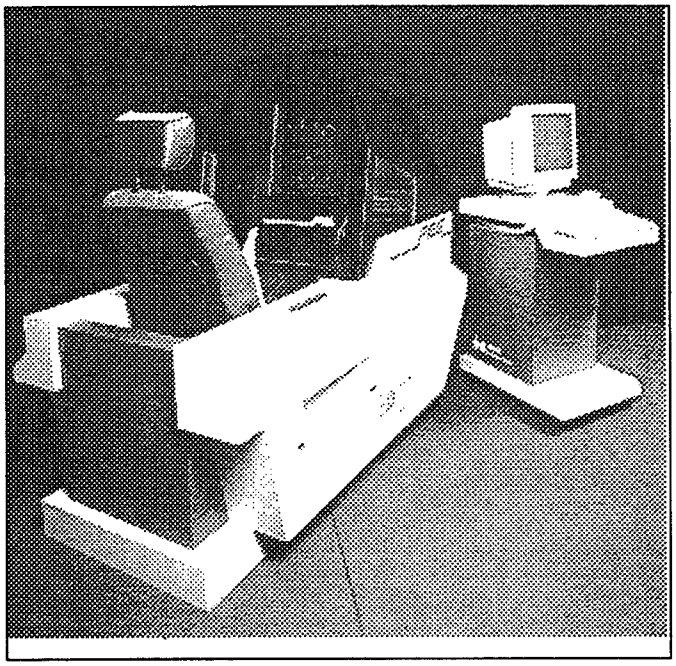


Figure 1. PES hardware.

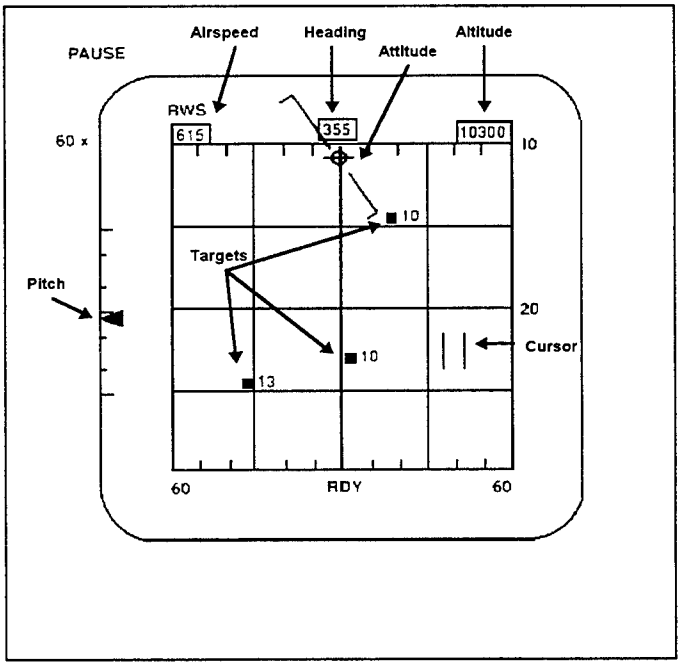


Figure 2. Typical Radar Screen.

disappear (see Figure 2 for a typical radar screen). When told to Match Target Altitude, the subject was required to maintain the same altitude as the target, using the stick. To Match Target Velocity, the subject was required to maintain the same velocity as the target, using the throttle and stick. To Match Target Heading, the subject was required to keep the target directly in the middle of the radar screen (straight ahead) using the stick. For the task of Tone Response, the PES would play either a high or low tone at random intervals through the subjects' headphones. The subject was required to respond to a low tone by pressing a button on the stick and to a high tone by pressing a button on the throttle.

PES Scoring. For each sub-task, the PES software automatically graded time and accuracy of responses. In addition, the software recorded false alarms (responses at inappropriate conditions). All scores were compiled into a composite PES score. The composite score was the score of the intercept task plus the scores of the secondary tasks; however, the secondary tasks were weighted twice that of the primary task.

Digit Task. The tape player was used to play random single digit numbers at approximately one second intervals. The subject was required to remember the last two numbers spoken and verbalize the difference between the two numbers.

NASA TLX Data Collection. The NASA TLX data were collected and evaluated in accordance with guidelines established by Hart and Staveland (1988). The data collection used a computer program to prompt the subject and record the data. The computer was situated within viewing range of the subject while seated in the PES. The subject rated each scenario using the TLX sub-scales: mental demand, physical demand, temporal demand, performance, effort, and frustration level. The sub-scales were generated by the computer to mimic the paper form of the TLX. Each task required the subject to rate each of six sub-scales. After the completion of all the tasks, the subject performed pairwise comparisons of each sub-scale that were used by the computer to calculate the overall workload rating for each scenario for each subject.

Procedure

Subjects were instructed to complete the training slide show before arriving at the test room. If the subject had not completed the training, they were allowed to view the presentation and ask questions. Next, the subject filled out the subject information on the top of the data collection sheet. Then, the subject flew four practice scenarios followed by nine experimental scenarios, which were counterbalanced using a modified Latin square. Table 1 shows the sub-tasks performed in each of the nine scenarios. Following each scenario, the subject completed the NASA TLX rating using the computer. Following all experimental trials, the subject completed the NASA TLX pairwise comparisons. Total experimental time was between 55 and 65 minutes.

Results

Four stepwise linear regression analyses were performed on the data. In the first, the NASA TLX sub-scale ratings were regressed onto total workload to determine the relative importance of each of the six factors on overall workload (see Table 2). The table shows that time demands, performance, and effort accounted for approximately 97% of the variance in overall workload.

Next, the overall PES performance was regressed on the number of concurrent tasks and found that the number of tasks accounted for 16.7% of the overall variance in performance, $F(1,214) = 42.85$, $p < .0001$. There was a correlation of -.314 ($p < .05$) between overall workload and overall PES performance. This relationship is shown in Figure 3.

Table 1. PES Scenario Tasks.

#	Tasks Performed	Total # of Tasks	Resources a/v	v/s	v/m
1	digits task	1	1/0	1/0	1/0
2	intercept target	1	0/1	0/1	0/1
3	digits, intercept, match target velocity	2	1/1	1/1	1/1
4	keep target in center, match target altitude and velocity	3	0/3	0/3	0/3
5	digits, tones, intercept, match target velocity	4	2/2	1/3	1/3
6	tones, intercept, match target altitude and velocity	4	1/3	1/3	1/3
7	digits, tones, intercept, match target altitude and velocity	5	2/3	1/4	1/4
8	tones, intercept, keep target centered, match altitude and velocity	5	1/4	0/5	0/5
9	digits, tones, intercept, center target, match altitude and velocity	6	2/4	1/5	1/5

Next, the overall workload ratings were regressed on to eight independent variables (see Table 3). Six of the variables represented the number of resources demanded for each task in each of the six possible "bottlenecks" of information processing. This analysis showed that the number of tasks was the most significant factor in overall workload. The number of tasks was then removed from the equation to examine the dimensions of the multiple resource theory. Table 4 shows that the two central processing resources accounted for nearly half of the variance in overall workload.

Discussion

The total number of concurrent tasks was an important determinant of workload and performance (see Figure 3). These results indicate subjective workload measures can be a robust predictor of performance, even in high workload situations. More interestingly, the total number of tasks was significantly correlated with the overall workload rating, accounting for 42.3% of the total variance. This statistic indicates that merely counting the number of concurrent tasks is a useful method of estimating overall workload. The number of tasks even accounted for 16.6% of the variance in PES performance. This result supports the Yeh and Wickens (1988) implication that the number of time-shared tasks would be a good predictor of workload.

Figure 3. The Effects of Number of Tasks.

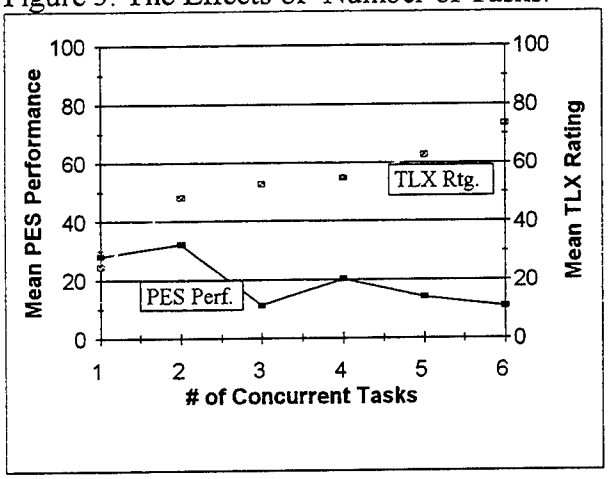


Table 2. Overall Workload Regression Analysis.

Step #	Variable	Cum R^2	Final Beta
1	Time Demands	.870	.933
2	Performance	.929	1.027
3	Effort	.969	1.066
4	Frustration	.979	1.093
5	Physical Demands	.989	1.132
6	Mental Demands	.995	1.151

Table 3. Overall Workload Regression Analysis.

Step #	Variable	Cum R^2	Final Beta
1	Total Number of Tasks	.423	.650
2	Verbal Central Processing	.460	.821
3	Auditory Input Channel	.476	.934

Table 4. Overall Workload Regression Analysis.

Step #	Variable	Cum R^2	Final Beta
1	Spatial Central Processing	.323	.568
2	Verbal Central Processing	.460	1.00
3	Visual Input Channel	.476	1.10

The present results also support the contention of Yeh and Wickens (1988) that the efficiency of time-sharing strategies has a large affect on workload ratings (see Table 2). The TLX "time" sub-scale accounted for 87.0% of the variance of overall workload, possibly because of the strict temporal demands of the digits task (which was not included in the Aretz et. al. study). However, the data did not support the idea that subject frustration would be the main factor affecting workload ratings when working memory is at full capacity.

The stepwise regression analysis presented in Table 4 has implications for multiple resource theory and subjective workload ratings. Spatial central processing demand was the variable that accounted for the most variance in the overall workload rating (of the six in Multiple Resource Theory), accounting for almost a third of the variance. Verbal processing accounted

for an additional 13.7% of the variance. The fact that the number of central processing demands (verbal and spatial) combined to account for approximately half of the variance in overall workload means that subjects' ratings were sensitive to working memory demands as predicted by Yeh and Wickens (1988).

Conclusions

The first important conclusion suggested by these data is that the NASA TLX is indeed sensitive to demands in high workload situations. Second, it seems that subject's perception of time pressure is a significant factor in subjective workload ratings, even when the subject is overwhelmed with tasks. Third, an important predictor of workload is the number of concurrent tasks, especially if these tasks demand central processing resources. This conclusion is an extension of Yeh and Wickens (1988) contention that time-sharing efficiency is an important dimension of subjective workload ratings. Future research should focus on developing a specific model that could be used to define workload using multiple resource theory.

References

Aretz, A. J., Shacklett, S., Acquaro, P., & Miller, D. (1995). The prediction of subjective pilot workload. In Proceedings of the 38th Annual Meeting of the Human Factors Society, pp. 94-97. Santa Monica, CA: Human Factors Society.

Hart, S. G. & Staveland, L. E. (1988). Development of NASA-TLX (Task-Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), Human Mental Workload, pp. 139-183. Amsterdam: Elsevier.

Hill, S. G., Iavecchia, H. P., Byers, J. C., Bittner, A. C. Jr., Zaklad, A. L., & Christ, R. E. (1992). Comparison of four subjective workload rating scales. Human Factors, 34, 429-439.

Wickens, C. D. (1992). Engineering Psychology and Human Performance. New York: Harper Collins.

Yeh, Y. & Wickens, C. D. (1988). Dissociation of Performance and Subjective Measures of Workload. Human Factors, 30, 111-120.

Responses of General Aviation Pilots to Autopilot and Pitch Trim Malfunctions

Dennis B. Beringer
Human Factors Research Laboratory
FAA Civil Aeromedical Institute, Oklahoma City

Abstract

A number of accidents and incidents have been traced to the interaction between the pilot and the onboard automated systems designed to reduce pilot workload and to decrease variability of aircraft performance. An examination of the use of the autopilot in simulated general-aviation flying was performed using 29 experienced pilots with complex aircraft time, 27 of whom had autopilot experience. Data collection was performed in the Civil Aeromedical Institute's Advanced General Aviation Research Simulator, configured as a Piper Malibu, for four simulated autopilot/pitch-trim failures. Detection/correction times and response strategies are discussed.

Introduction

The most visible and recollected aircraft accidents are those which result in the loss of large com-mercial aircraft, such as China Airlines' Flight 140, April 26, 1994, on approach to Nagoya/Komaki airport, Nagoya, Japan (Katz, 1995). The data indicated that the aircraft, an Airbus A-300-600R, ultimately stalled and crashed after attaining a pitch-up attitude of approximately 52 degrees at 78 knots. The problem appeared to be the pilot's continued attempts to fly the airplane manually with the auto-pilot engaged in go-around mode. The captain, who had apparently inherited the approach from the first officer after an autothrottle, but not autopilot disengagement, ultimately lost the struggle with the aircraft as the autopilot trimmed the aircraft nose up after the captain's continued attempts to force the nose down. Problems with automated systems are not restricted to commercial carriers. Similar incidents involving pitch trim malfunctions and other autopilot difficulties have been reported for general aviation (GA) aircraft (Wilson, 1995; Katz, 1995).

Present certification standards require that an autopilot system, in a hard-over failure where the control surface servo is driven at its maximum rate, cannot place the aircraft in greater than a 60-degree bank nor place undue loads (0 - 2 Gs limits) on the airframe "within a reasonable period of time" (F.A.R. 23.1329). This has been operationalized (Advisory Circular 23.1329-2, Automatic Pilot System Installation in Part 23 Airplanes, 3/4/'91) as within the three seconds following the initial detection of the uncommanded bank. Similarly, this restriction applies to pitch and pitch trim tests to the degree that the aircraft cannot stall, exceed limit speeds, or require excessive control force during recovery at the end of the three-second period. This time interval supposedly provides three seconds in which the pilot can diagnose the problem and take corrective action (autopilot disconnect is assumed). A delay of one second was adopted for malfunctions on a coupled approach on the theory that the pilot is likely to be attending to the instruments more closely on approach than during cruise. Cooling and Herbers (1983) noted, in

their discussion of human factors, that "...there are no studies available to support the FAA certification standard of a three second delay (enroute) or a one second delay (on approach) before initiation of recovery by the pilot from an autopilot malfunction." How-ever, it has been suggested that the data were actually derived from a study of airline pilots' responses collected during a study performed at Wright-Patterson AFB in the 1960's (ACE-100, 1996). The focus of our research, in support of Aircraft Certification, was the responses of pilots to overt and subtle autopilot malfunctions and the factors influencing the speed and the selection of those pilot responses.

Method

Design/Subjects

The experimental approach, a single-factor within-subject design using autopilot malfunction type (4) as the independent variable, was selected because high between-subject variability in response times to the malfunctions was expected. The four malfunction types were: "command over" roll failure (rate = 6 deg/sec), soft roll failure (rate = 1 deg/sec), soft pitch failure (rate = 0.2 deg/sec), and runaway pitch trim. Dependent variables recorded included flight performance variables and states of critical switches; autopilot disconnect, engage, circuit breaker, and pitch trim switches and circuit breaker. Pilots were obtained from the local area who were instrument rated and had experience with complex aircraft and autopilot systems. Age ranged from 24 to 72 years (median = 42) and the sample contained 27 men and 2 women. None had less than 300 hours of flight experience.

Equipment/Procedures/Tasks

Data collection sessions were conducted in the Advanced General Aviation Research Simulator (AGARS) in the Human Factors Research Laboratory, Civil Aeromedical Institute. The simulator was configured as a Piper Malibu with Bendix/King avionics (KFC-150 autopilot); software approximated behavior of both but exact flight equations were not available. High-fidelity primary flight displays were presented in the cockpit on three masked CRTs that replicated the Malibu panel layout and gave the appearance of hard, dedicated instrumentation. The out-the-window depiction spanned 150 degrees of visual arc and was a high-resolution textured representation of the Oklahoma City area.

Pilots participated in one 2- to 2.5-hour session. Pilots were told that the study was to examine how pilots used autopilots in routine flying and to gather opinion data on useful features. The first hour consisted of experiment-related paperwork and familiarization training activities, including: reading excerpts from the autopilot manual, cockpit familiarization, and a half-hour familiarization flight using all autopilot modes. The second half of the session was used to collect performance data for the four malfunction conditions. A simple round-robin instrument clearance was flown from OKC to two local VOR stations and back in IFR conditions between textured cloud layers (distinct visual horizon but no ground detail), requiring pilots to interact with ATC, fly vectors, track inbound to two VOR stations, and fly a fully-coupled ILS approach. Pilots were instructed to fly as much of the course on autopilot as possible. Malfunctions were spaced such that sufficient time elapsed between failures (13-15 minutes) to prevent interference between episodes. Command roll and soft pitch were encountered in level flight, soft roll during descent,

and half pitch trim during the ILS approach and half during ascent from 6000' to 7000'. Data collection flights averaged approximately 1.2 hours. The session concluded with an autopilot-experience questionnaire and interview to determine each pilot's depth of knowledge of autopilot and autotrim malfunction consequences. Only the pitch trim malfunction produced both auditory and visual warnings.

Results

Response Times

Command roll (roll servo). Of all the failures, commanded-roll and pitch-trim failures were rated as easiest to diagnose (11 of 26 votes for each). The commanded-roll failure emulated an autopilot-commanded roll that failed to stop at the target bank angle. Preliminary analyses for both roll malfunctions and the soft-pitch malfunction are based upon time from initial failure to disconnect of the autopilot by any means (yoke-mounted disconnect, panel disengage, circuit breaker). Times ranged from 1.78 seconds to 124.11 seconds (Mean = 20.5; Median = 8.77). However, 69 % of the pilots disconnected within 13 seconds of the initial failure and half within 8 seconds. These "immediate" disconnects by 15 of the 29 pilots were defined by sequences where no other significant actions occurred between failure onset and autopilot disconnect. Using a response time of 8.7 seconds or less as a cutoff value, 93.7% of the sample of "immediate" responders was included. Thirteen pilots chose to manually override the autopilot, whether by using the control-wheel steering option or by simply overpowering the roll servo, without disconnecting the autopilot, with 90% having response times of 48.3 seconds or less. A post-hoc comparison of the log-transformed disconnect times of the two groups, with the highest and lowest extreme times removed, indicated a significant difference ($F[1,24] = 53.27$, $p<0.0001$) between the immediate disconnects (untransformed mean = 5.93 seconds) and the manual overrides (untransformed mean = 28.26 seconds). Distributions are shown in Figure 1.

Soft roll (roll sensor). The soft-roll failure was rated as third in difficulty to diagnose, but was rated easiest to correct (13 of 26 votes). Following removal of one outlier (194 seconds), pilot performance was again categorized as immediate disconnect (16) or manual override (12). Those categorized as immediate disconnect responses averaged 11.72 seconds (range: 4.52 to 16.69) while those categorized as manual overrides averaged 37.45 seconds (range 13.16 to 85.14). Approximately 88% of all immediate disconnects occurred in less than 17 seconds, with 75% occurring in less than 14 seconds. Post-hoc comparison indicated the mean difference to be significant for both raw and log transformed scores (log scores: $F[1,26] = 27.07$, $p<.00005$).

Soft pitch (pitch sensor). The soft-pitch failure was rated as most difficult to diagnose (12 of 26 votes), and was rated third easiest to correct, missing a tie for second by one vote. Performances were again categorized as either immediate disconnect (12) or manual override (17). Three pilots never diagnosed the failures, manually flying the airplane without disconnecting the autopilot, and their scores and one other outlier were removed, leaving 13.

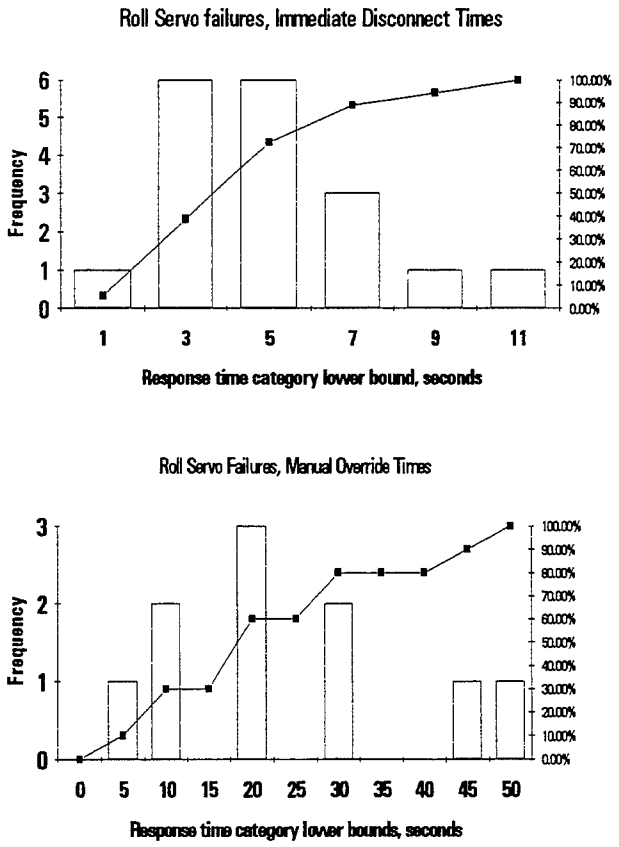


Figure 1. Response time distributions and cumulative frequency plots, commanded-roll failure, for immediate disconnects and manual overrides.

Immediate disconnects averaged 17.38 seconds (range: 6.5 to 31.5) and manual overrides averaged 46.19 (range: 15.2 to 76.2). Approximately 50% of immediate disconnects occurred in less than 16 seconds, with approximately 85% occurring in less than 24 seconds. Post hoc comparison of the log-transformed data showed the distributions of the two types of responses to be significantly different ($F[1,22] = 20.69$, $p<.0005$).

Runaway pitch trim. This failure was different from the others in that only the Pitch Trim circuit breaker would correct the problem. The interim solution was the AP disconnect/trim interrupt switch. Only three pilots chose the optimal response of depressing and holding the disconnect switch followed by pulling the circuit breaker. Four others depressed and held the disconnect switch at various times in the recovery. The vast majority of initial responses were yoke AP disconnect (15), followed in frequency by panel-mounted AP engage switch (5), mode manipulation (2), manual override (2), and the pitch trim circuit breaker (1). Overall, 21 of the 25 pilots considered were classified as "immediate" responders, two were classified as manual overriders, and two as mode changers.

Two stages of response were of interest; first, the time required to detect a malfunction and initiate some action (autopilot disconnect, control-wheel steering, autopilot engage or circuit breaker) and second, the time lag between the initial action and the pulling of the pitch-trim circuit breaker. Average time to initial action for the usable 25 pilots was 10.46 seconds, with all except one response over 3 seconds. One can see in Figure 2 that 50% of the responses occurred in less than 7 seconds, with 65% of the cases in less than 9 seconds. Latencies to the pulling of the pitch trim circuit breaker averaged 35.4 seconds (range: 4.91 to 109.69),with an average lag of 22.69 seconds (high and low scores removed) between the initial response to the runaway pitch trim and the final remedy.

Initial examination of the questionnaire and interview data indicated that all pilots understood that they could manually overpower the autopilot servos, and that 22 were aware of the potential interaction between a runaway pitch-trim motor and autopilot pitch-attitude (elevator servo) inputs. Another 4 had not considered the potential interaction previously, but grasped the concept immediately during the interview. When asked what their strategy for dealing with autopilot malfunctions was, the group voiced two anchor strategies and a combination of the two as a third. The immediate-disconnect strategy was endorsed by nine individuals, while two others expressed a procedural approach that was closely related to the immediate disconnect strategy. Another five individuals suggested that they would fly the aircraft through the malfunction while attempting to diagnose the problem. A third group took a middle-of-the-road stance, saying that the strategy was malfunction dependent. These seven expressed their strategies as, "Fly through mild failures; disconnect for severe failures," or "diagnose while the unit is still engaged, then disconnect."

Discussion/Conclusions

Present certification practices assume a malfunction will be either severe enough to produce supra-threshold cues or that some type of alerting mechanism will warn the pilot of the autopilot malfunction, thus starting the three-second "recognition" period. Flight test personnel (FAA Aircraft Certification Service, 1996) have reported instances where malfunctions have gone undetected until the test administrator or safety pilot pointed them out, sometimes after criterion limits had been reached. In these cases, the autopilot failed to obtain certification. Our data indicate that pilots responding to a clearly supra-threshold failure, the commanded roll failure, and who are intent upon an immediate response require an average of 5.93 seconds to respond with an autopilot disconnect, some requiring as long as 11.8 seconds. In fact, only one of the 18 pilots classified as immediate responders acted within three seconds. It is general practice, for "obvious" malfunctions, to allow one second for detection, producing a four-second interval within which the pilot is to both detect and respond to the malfunction, almost two seconds shorter than the mean sample response. In the case of the commanded roll failure, one could accommodate 90% of the present pilot sample by specifying 9 seconds as the upper bound of the interval. A more conservative approach, using 7 seconds as the criterion, would still account for the responses of over 70% of the sample. One should note that at the usual 5 deg/sec

commanded roll rate, a 60-degree bank would not be exceeded for 12 seconds. However, a roll-servo hard failure, at approximately 15 deg/sec for this class of aircraft, would do so in four seconds. Flight attitude data are presently being reduced to determine maximum deviations in pitch, bank, airspeed, and altitude, but an initial examination indicated that very few, if any, pilots exceeded any limits during recovery.

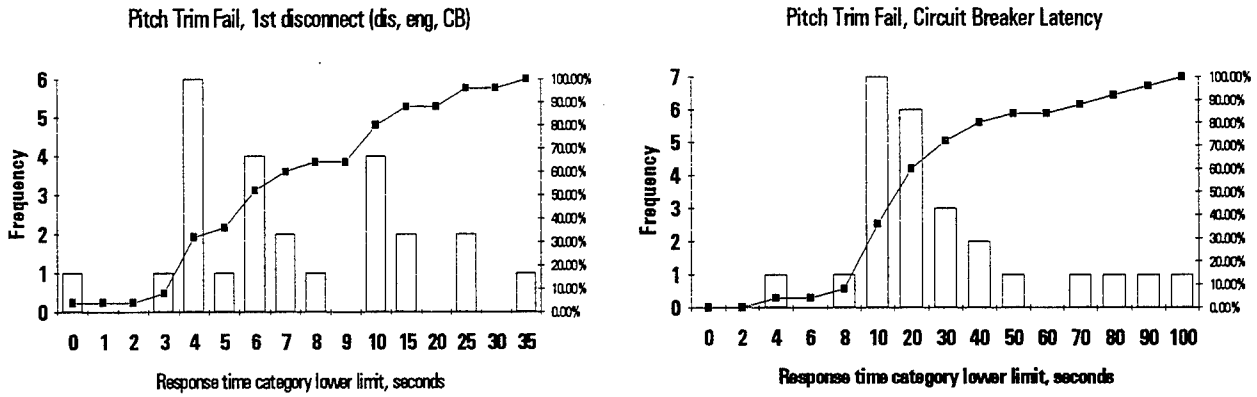


Figure 2. Frequency distributions and cumulative frequency plots of first disconnect and pitch-trim circuit breaker response times.

It was not surprising, given the comparatively low rates of change in the more subtle failures, that significantly longer intervals were required for pilot response (roll, 11.72 seconds: pitch, 17.38 seconds). Because the attitude indicator (ADI) continued to accurately depict attitude during these malfunctions (true sensor failure would not), detection times were probably shorter than would otherwise be expected. Average first response to runaway pitch trim was 10.46 seconds, no doubt contributed to by the fact that the simulated system did not immediately disconnect upon occurrence of the runaway as would the actual autopilot. This allowed the pitch servo to compensate for (and mask) the initial trim deflection. The auditory trim malfunction warning did, however, provide an immediate cue.

It is probably safe to assume that for failures accompanied by high acceleration rates, the present guidelines are adequate when the required response is simple. The application of the present guidelines to the more subtle malfunctions depends upon aircraft performance limits not being exceeded by "detection plus three" and, thus, upon the pilot ultimately detecting the malfunction either unaided or with the assistance of a warning device. By the present standards, a subtle malfunction that places the aircraft in an unacceptable attitude without the pilot's detection disqualifies that autopilot system. The currently available countermeasures are either to design the system so that it does not exhibit "soft" failure modes (extremely high reliability, e.g., multiple attitude sensor sources), or to annunciate all detectable failures with visual/auditory signals.

Acknowledgments

The author thanks Mr. Barry Runnels for simulator engineering support during the course of the study and Mr. Howard Harris for his efforts during data collection and data reduction. FAA Aircraft Certification Service (AIR-3) sponsored the research; the Small Airplane Directorate (ACE-100) coordinated the research.

References

ACE-110 (Small Airplane Directorate, FAA) (1996). Personal communcation.

Cooling, J. E. & Herbers, P. V. (1983). Considerations in autopilot litigation. J. of Air Law & Commerce, 48, 693-723.

FAA Aircraft Certification Service (1996). Personal communication with flight test personnel.

Katz, P. (1995). NTSB Debriefer: The dark side of "Otto pilot". In Plane & Pilot, (February), 31(2), 18-19.

Wilson, B. G. (1995). I learned about flying from that (#660): Unacquainted with the autopilot. In Flying, (June), 122(6), 122.

The Opto-kinetic Cervical Reflex (OKCR) in Pilots of High-Performance Aircraft

Ronald F. K. Merryman, M.S.
United States Air Force Academy
Anthony J. Cacioppo, Ph.D.
Wright State University

Abstract

Background: For over sixty years, researchers and engineers have based investigations and the design of cockpit displays and structures upon the presupposition that, during flight, the pilot maintains a head alignment coincident with the aircraft's vertical axis (z-axis). Recent simulator studies have verified the existence of a pilot neck reflex which refutes this long-standing assumption. This reflex, named the opto-kinetic cervical reflex (OKCR), occurs during visual flight and is theorized to be an attempt by the pilot to stabilize a retinal image of the horizon to maintain spatial orientation. As a result, during initial banking maneuvers, pilots view a fixed-horizon image and not a moving-horizon. The research objectives were to determine if the OKCR occurs during actual flight of high performance jet aircraft and to model the response. Hypothesis: Pilots of high performance aircraft will exhibit the OKCR. Additionally, the OKCR is dependent on the phase of banking (entering into or exiting from a banked position). Methods: This was an observational study in which the head positions of nine pilots were recorded during actual F-15 aircraft flight and subsequently analyzed. Results : Objective data indicate the OKCR caused pilots to tilt their heads during aircraft bank ($p < 0.0001$). Also, the reflex was found to be independent of the bank phase. Conclusion: The OKCR was shown to be a strong, natural response and the flight results correlated extremely well with simulator results. The impact of these results on pilot training, spatial disorientation, physiological injury and safety, and the re-design of displays for aircraft attitude and virtual reality are discussed.

Recent flight simulator studies into pilot head orientation during flight have refuted a previous, long-standing assumption that pilots always align their head and body vertically with the aircraft (in the Z-axis) throughout all flight maneuvers. This original premise was stated in 1936 following the successful implementation and employment of the first attitude indicator display. The statement was merely an educated observation and was never supported through scientific evidence or testing.

The persistence of the assumption for the past six decades has been due to a number of issues. Primarily, the assumption had never been challenged by actual scientific studies involving pilot head alignment. Secondly, the assumption had been propagated via pilot training and education which discourage motion of the head during flight. This training leads pilots to "believe" they do not tilt their heads during flight. Finally, no direct link had been established between aircraft mishaps, aircraft displays and the possibility that the1936 statement was incorrect. The closest attribution to that link is that of spatial disorientation (SD). SD is typically attributed as a cause, in and of itself, of human error (and mishaps) but not as a result from a conflict between aircraft displays and the reality of pilot head alignment during flight.

Two recent investigations (Patterson, 1995 and Smith, 1994) have documented the existence of a pilot reflex currently named the opto-kinetic cervical reflex (previously: opto-kinetic collic reflex), or OKCR (Patterson, 1995). Both studies have found that pilots naturally tilt their heads during aircraft bank in an apparent attempt to align their eyes with the visible horizon. This reflex occurs during visual flight but not during instrument (no external visual stimuli) flight. The discovery of the OKCR is important since pilots have been trained to minimize their head motion during flight. Both studies were completed in non-motion aircraft simulators. Until this time, no studies have objectively investigated the existence of the OKCR during actual flight. The purpose of our research was to determine if the opto-kinetic cervical reflex occurs during actual flight of high performance jet aircraft. The focus was on the lateral flexion reflex (angle of head tilt [left and right]) in response to aircraft bank (or roll) angle. The connection between these variables will provide information as to which environmental sensory cues are important to the pilot in order to maintain the aircraft's attitude.

Method

Subjects

Nine USAF operational fighter test pilots participated. Each volunteer was male with six to twelve years experience flying high-performance fighter aircraft and currently flying the F-15 aircraft. All pilots were instrument qualified. Pilots were advised that the purpose of the study was to evaluate normal pilot reflexive actions during various phases of flight. This was a blind investigation and therefore the pilots were not be briefed on the actual variables until the completion of the study.

Apparatus

The subjects piloted McDonnell-Douglas F-15C fighter aircraft based at Nellis Air Force Base, Nevada. As this investigation involved actual flight in high-performance fighter aircraft, it was imperative that non-invasive methods of data collection were employed. Two F-15 aircraft were equipped with Polhemus® MAGNETRAK magnetic head tracker systems which allowed the collection of pilot's head motion parameters without interfering with the pilot's tasks. This satisfied the requirement for passive, non-invasive data collection.

Procedure

This was an observational study and therefore no experimental task was designed. All data were collected from normal, day-to-day aircraft sorties and missions flown at a Nellis Air Force Base. All missions occurred during VMC (visual meteorological conditions) flight. These sorties and missions were not specific to this study. Pilot subjects flew sorties during which various maneuvers and engagements took place. Both aircraft position data and pilot head orientation data were simultaneously recorded via telemetry during the flights.

Data Collection

During flight, pilot head orientation and aircraft dynamic parameters were continuously sent from the aircraft to ground stations via near-real-time electronic telemetry signals. These parameters were then stored as raw data on magnetic tapes for each pilot and mission. The data sampling rate was ~10 samples per second (or approximately one data point every 100 milliseconds). Data reduction was used to create blocks of data for which only the independent and dependent aircraft and head orientation parameters of interest were retained.

Variables

The dependent variable, ROLL, was the pilot head tilt angle as measured from body vertical; negative ROLL values corresponded to a lateral flexion tilt to the left. The two independent variables were BANK and PHASE. BANK was the aircraft angle of bank with respect to the Earth's horizon; negative BANK values corresponded to a left aircraft bank. The second independent variable, PHASE, was a qualitative variable with two values: INTO and OUT_OF. To investigate if subjects' head tilt response may have been dependent upon the phase of the aircraft turn, the data was divided into two categories: head tilt while entering (INTO) the banked turn and head tilt while exiting (OUT_OF) the turn. The nine pilots in the study were considered to be a random sample from the population of possible pilots. The aircraft bank angle, broken down by 5° levels were considered of interest in themselves and were therefore fixed.

Results

Subject Data

All nine subjects completed a sufficient number of maneuvers (aircraft banking turns) to provide a quantity of raw data equivalent to or greater than that used in the simulator studies. Data were converted into a 2 x 37 matrix for each pilot. These matrices were used for analysis. The mean head tilt for each pilot at each aircraft bank angle was the dependent variable in the matrices for two reasons: 1) this method provided a balanced ANOVA approach via one head ROLL observation per aircraft BANK angle, and 2) this was the method used to analyze the simulator study results.

Interaction and Main Effects

The main effect (Table 1) for PHASE of the aircraft turn, characterized by the increasing or decreasing angle of aircraft bank, was not found to be statistically significant ($F(1,8) = 5.3176$, $p = 0.7169$). Furthermore, there was no significant interaction between BANK and PHASE. Therefore, data were pooled leaving a single factor, repeated measures model design (Table 2). There was a significant effect of aircraft BANK angle upon the subjects' head ROLL angle: ($F(36,325) = 1.4534$, $p < 0.0001$).

Regression Analysis

Data from simulator studies suggested the OKCR response is sigmoidal in shape with a linear phase between ~±45° at which point it levels-off asymptotically. Trend analysis was used to determine the components of the model. Initially four forms were tested: linear, quadratic, cubic, and quartic. Following the significant results of aircraft BANK upon the pilot head ROLL angle (tilt of head), a regression procedure was used to determine the coefficients of the response. As predicted, the linear and cubic parameters were found to be statistically significant ($p = 0.0002$ and $p = 0.0013$, respectively), while the quadratic and quartic components were not statistically significant ($p = 0.1550$ and $p = 0.0992$, respectively). These results were produced via the POLYNOMIAL option in the SAS GLM procedure. The equation used to fit the model was:

$$ROLL = \beta_0 + \beta_1 \times BANK + \beta_2 \times BANK^2 + \beta_3 \times BANK^3$$

The results of the regression procedure are in Table 3 (parameter estimates). Figure 1 shows the plot of the predicted polynomial response based on the regression analysis. The model is indicated by the solid line with triangle markers; the overall mean of individual pilot responses are annotated by the plot of open squares.

Source	df	MS	F	p
PHASE	1	5.6883	0.14	0.7169
Subject x PHASE	8	40.3068		
BANK	36	15133.4431	15.43	0.0001
Subject x BANK	288	980.6711		
PHASE x BANK	36	106.5359	1.22	0.1924
Subject x PHASE x BANK	288	87.5298		

Table 1. Two-Factor ANOVA Results

Source	df	MS	F	p
BANK	36	2005.1539	21.78	0.0001
Subject x BANK	325	92.0757		

Table 2. Single-Factor ANOVA Results

Variable	df	Parameter Estimate	Standard Error	T for H_0: $\beta_x = 0$	Prob > |T|
INTERCEPT	1	-0.440250	0.73632766	-0.598	0.5540
BANK	1	-0.399810	0.02303284	-15.622	0.0001
BANK^2	1	0.000419	0.0001926	2.176	0.0368
BANK^3	1	0.000016988	0.00000412	4.122	0.0002

Table 3. Regression Analysis Parameters

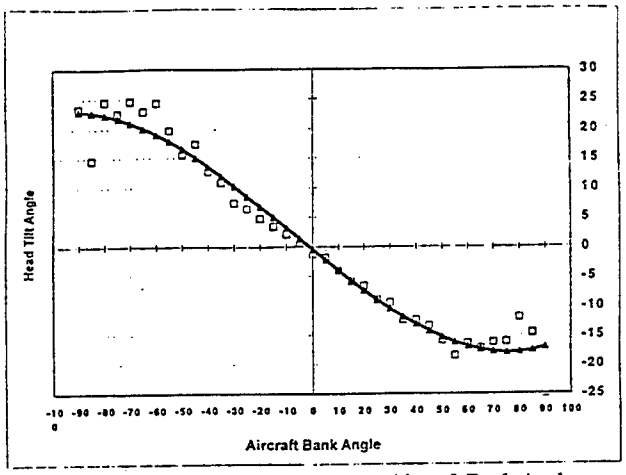


Figure 1. OKCR: Head Tilt vs. Aircraft Bank Angle

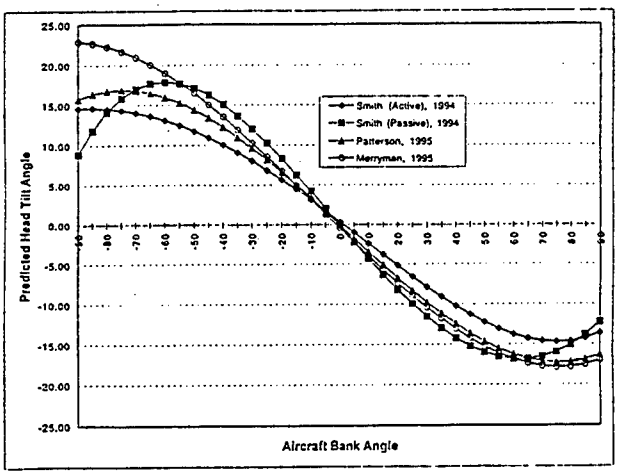


Figure 2. Four OKCR Models

Discussion

The results of this study indicate that the opto-kinetic cervical reflex is an irrefutable behavior of pilots in high performance jet aircraft. This objectively confirms the subjective observations as well as validates the work completed in the simulator studies. Figure 2 shows a plot of the four OKCR models for comparison. Each line is a plot of head angle versus aircraft bank angle. The four models are: this study's third-order model, Patterson's (1994) third-order model and Smith's (1994) active and passive fourth-order models. Graphical inspection indicates a very good match between all four models. To compare the actual flight data against the simulator models, the method of standardized residuals was utilized. Each subject's mean head tilt response at every aircraft bank angle (from actual flight data) was compared with the predicted head tilt from the (simulator) models. A minimum of 95% of all the standardized residuals fell within the normal range for each of the three models considered. Therefore, the OKCR flight data was found to be statistically comparable to the results from the previous simulator studies.

Despite the many confounding factors possible in an observational study such as this, the OKCR was significant enough to overcome these extraneous variables. An approximation of the simulator studies results was predicted, but the actual level of coincidence between the three studies was extremely surprising. The fact that the OKCR can be induced in a motionless simulator, without the true physical and vestibular effects of actual flight, also suggests that the reflex is a powerful, natural behavior based primarily on visual inputs.

Conclusions

This investigation verified that the opto-kinetic cervical reflex does occur in the cockpit of high performance aircraft. It is a seemingly natural response to a very unnatural stimulus: rotation in the roll (coronal) plane during airborne motion. As Patterson (1995) stated, it is theorized that this response is an attempt to stabilize the retinal image of the visible horizon. The stabilized image becomes the primary source of visual information used to maintain spatial orientation. OKCR is a logical reflex considering, to a pilot flying in an aircraft, there is only one physical visual stimulus which can be used to determine body orientation: the Earth. And the best discriminator on the earth's surface is the horizon, the natural divider between the ground and the sky - the pilot's medium. Therefore, the pilot reflexively seeks to maintain a relatively fixed head-horizon orientation as long as possible. While this accounts for spatial orientation of the human portion of the system, the pilot is still attached to an aircraft. Keeping the aircraft from impacting the ground is a prime concern for the pilot and therefore the pilot must also account for the spatial orientation of the airframe with respect to the earth. To accomplish this, Patterson (1994) has suggested that the aircraft wing tips (and other aircraft structures) act as peripherally viewed secondary sources of information by which pilots detect the independent motion of the aircraft relative to their own head.

In summary, during low angles of bank (AOB) pilots maintain a head-horizon orientation from which spatial awareness is determined. Once the maximum OKCR head tilt is exceeded (corresponding to high AOB), the pilot's head becomes "attached" to the pilot's body and

aircraft. The complete system is now rotating during the aircraft bank. The pilot is now maintaining a head-aircraft orientation. When returning to a "wing's level" attitude, the pilot maintains a head-aircraft orientation until the aircraft AOB is about ±45° at which the pilot seeks a head-horizon orientation. The transition between head-aircraft and head-horizon orientations represents a critical change in the pilot's cognitive view of the world since the frame of reference changes instantaneously.

References

Merryman, R.F.K. (1995). The Opto-kinetic cervico reflex in high-performance aircraft. (Thesis) Dayton, OH: Wright State University.

Patterson, F.R. (1995). Aviation Spatial Orientation in Relationship to Head Position and Attitude Interpretation. Dissertation. Dayton, OH: Wright State University.

Poppen, J.R. (1936). Equilibratory functions in instrument flying. The Journal of Aviation Medicine, 6: 148-160.

Smith, D.R. (1994). Aviation Spatial Orientation in Relationship to Head Position, Attitude Interpretation, and Control. (Thesis). Dayton, OH: Wright State University.

Single Seat Fighter Pilot Landing Performance During Multiple, Long-Duration Missions

Patrick E. Poole, 1Lt, USAF
Daniel H. Bauer
Armstrong Laboratory
Brooks AFB TX 78235

Kory G. Cornum, Lt Col, USAF, MC
Wilford Hall Medical Center/Orthopedic Department
Lackland AFB TX 78236

Abstract

The impact of fatigue during two week-long schedules, respectively driven by different deployment times (0900 hrs and 2100 hrs), was objectively evaluated by measuring flying performance during simulated landings. Eleven operational F-16 pilots completed a nine hour simulated eastward deployment, followed by six simulated engagement sorties, resulting in a total of seven landings each. The missions were conducted in two full visual, non-motion based flight simulators. Results indicated no significant differences in landing performance between the two schedules. These findings do not support the hypothesis that a nighttime deployment (2100 hrs) would cause greater fatigue than a daytime deployment, and would increase cumulative fatigue effects during subsequent sorties. However, the results show the pilots landed equally well during the daylight or at nighttime. Rules for deployment are discussed in consideration of these findings.

Throughout history, combat operations and fatigue have been synonymous. People are pushed to their limits to achieve victory, and air combat is no exception. Desert Storm necessitated the United States Air Force (USAF) to more fully address the issue of fatigue, as pilots flew more sorties, more hours, and in a more demanding environment than in any previous conflict. This was especially true for single seat fighter pilots, who flew more hours in the first two weeks of the war than they would normally fly in five months of peacetime operations. General John M. Loh (ret.), former Commander, Air Combat Command, succinctly described the USAF's growing concern with fatigue in his letter to General Ronald W. Yates (ret.), former Commander, Air Force Materiel Command:

> For years we have "brute forced" our way through long deployments with suboptimally rested pilots and have been successful. This tempts us to believe "we can keep things as they are." However, as we invest further in night operations, understanding fatigue and circadian rhythm issues takes on greater importance. We need to better understand the implications of all types of fatigue (especially from inadequate rest and circadian disturbances) for deployments and extended operations. Line commanders must understand the impact of fatigue on safety and mission accomplishment.

With the end of the Cold War, the probability of two or more separate theater conflicts has increased dramatically, and the USAF has developed contingency plans accordingly. This possible division of force, coupled with the present reduction in force structure, could further exacerbate problems associated with fatigue. In addition, recent fatigue-related aircraft mishaps have highlighted this problem. The primary purpose of this study was to compare pilot landing performance under two schedules respectively driven by different deployment times.

Method

Subjects

The participants were eleven volunteer, operational F-16 pilots from Air Force regular, reserve, and guard units. Five of the pilots performed both schedules in the study (Table 1), for a total of 16 "subject runs". The pilots for each week's 2-ship formation came from the same unit, and therefore flew with one another on a regular basis. Age ranged from 30 to 43 years ($\underline{M}$ = 34.6 years), and pilots were within the military ranks of Captain to Lieutenant Colonel. The subject pool included seven instructor pilots, eight mission commander rated pilots and four Fighter Weapon Instruction Course (FWIC) graduates (Air Force program similar to Navy's Top Gun). Experience in the F-16 ranged from 600 to 2450 flying hours ($\underline{M}$ = 1477 hours), and deployment experience ranged from zero to eight overseas deployments ($\underline{M}$ = 3.7). The demographic data indicated a highly experienced subject pool. Subject pairs were randomly assigned to one of the schedules.

Table 1

Definition of Schedules

	Schedule A (n=8)		Schedule B (n=8)	
Sortie	Take-off/Landing Times *	Lighting cond. in sim. at landing	Take-off/Landing Times *	Lighting cond. in sim. at landing
Deployment	2100/0600	Daylight	0900/1800	Nighttime
Employment 1	2000/0300	Daylight	0800/1500	Nighttime
Employment 2	0900/1500	Nighttime	2100/0300	Daylight
Employment 3	2100/0100	Daylight	0900/1300	Nighttime
Employment 4	0400/1000	Daylight	1600/2200	Nighttime
Employment 5	1200/1400	Nighttime	0000/0200	Daylight
Employment 6	1600/2200	Nighttime	0400/1000	Daylight

Note: Each schedule began on Tuesday, at the respective time for the deployment. * All times are Mesa local time. Employment sorties included take-offs, aerial refuelings, combat air patrol, aerial engagements, and landings.

Apparatus

All apparatus are owned and operated by the Aircrew Training Research Division of Armstrong Laboratory (AL/HRA), located in Mesa, AZ. The cockpits used for the study were F-16 Multi-Task Trainers (MTTs). Each MTT is a high fidelity, F-16C cockpit incorporating all principle instruments, switches, flight controls, sensors, and offensive and defensive weapons systems. One MTT was placed in the Display for Advanced Research and Training (DART) while the second was stationed in the mini-DART. The DART consists of nine cathode ray tube (CRT) projectors which are rear-screen projected in a wrap-around-the-cockpit configuration. The real image, located 3.5 feet from the pilot's eye, has a peak luminance of 25 footlamberts (fl) at a contrast ratio of 50:1. The resolution is 4.25 arc minutes/pixel in a field of regard (FOR) as large as that available in an actual F-16C cockpit. A Polhemus head-tracker determines where imagery cannot be seen by the pilot, so that six image generator channels can be switched to cover the nine display projectors. Heads-Up-Display (HUD) imagery is provided with a separate higher-resolution projection on the front window. The mini-DART is a smaller version of the DART, and utilizes eight CRT projectors and four image generator channels. The wing man was placed in the DART for the duration of the study, while the lead "flew" the mini-DART.

Procedure

At the beginning of each week of data collection, the pilots were given an initial in-briefing, describing the purpose of the study, the apparatus to be used, rules of engagement, and the grading criteria. The pilots were instructed to "fly" the simulators as they would an actual aircraft in a real-world situation. Each week's pair completed one of the schedules outlined in Table 1. Dynamic gradual ambient light transitions were simulated throughout the deployment and engagement sorties and adverse weather conditions were simulated during take-offs and landings. Additional realism was introduced into the scenario by asking the pilots to use "piddle-packs" to urinate and eat within the MTTs during the sorties. Pilots were housed locally within the simulator building during the engagement sorties, and lighting and other environmental variables were controlled to the extent practical to simulate the actual time shift of a European deployment. Specifically for the landing, the pilots were tasked to fly a predefined instrument approach pattern to the simulated airfield (Figure 1), during which they were required to intercept precise checkpoints in terms of bearing, altitude, and range.

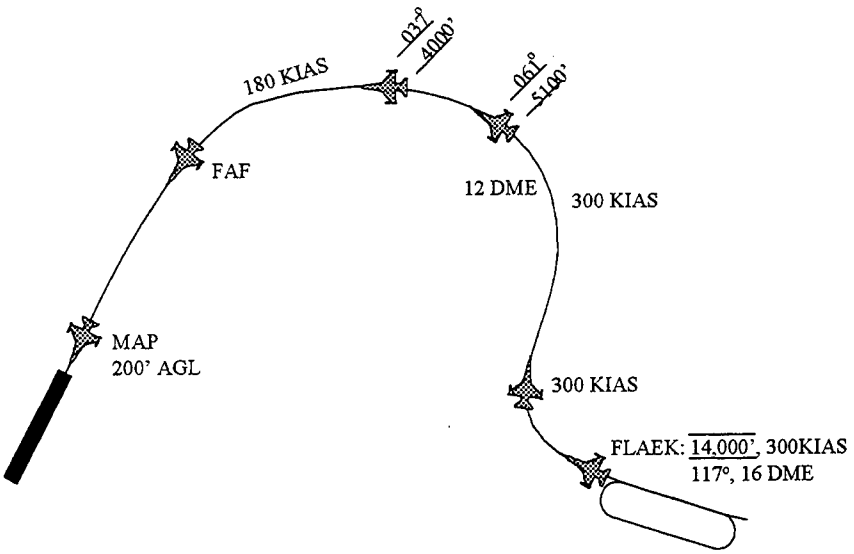


Figure 1. Graphical representation of approach pattern.

Scoring

The flight simulators recorded specific digital data from each workstation in the study, which allowed for objective, computer evaluated grading of the recovery/landing. For the recovery portion, deviations from three "hard" altitudes (represented in Figure 1 by bars above and below the altitude, e.g. $\overline{14{,}000'}$) were measured at the Initial Approach Fix (IAF) (FLAEK), and at the 061^0 and 037^0 radials. Also, the pilot's ability to maintain a 12 DME (Distance Measuring Equipment) arc from the TACAN (Tactical Air Navigation) at the airfield was measured between intersection of the arc and the 037^0 radial. Finally, during the final approach portion (between the Final Approach Fix (FAF) and the Missed Approach Point (MAP)), vertical (glideslope) and horizontal (localizer) deviations from an optimal ILS (Instrument Landing System) approach were measured.

Results

The study evaluated the objective scores of six variables involved in the simulated landing. Because of the administration of a performance maintenance medication (dextroamphetamine) to certain pilots during the seventh mission, the landing scores for this sortie were not considered in the evaluation. Root mean square (RMS) error scores were calculated for the 12 DME arc, glideslope, and localizer deviations, and the absolute value was taken of all altitude deviations. Table 2 shows the means for all measured variables. An analysis of variance of the means was performed, which looked at landing performance differences between the two schedules, as well as flying conditions at landing (daylight vs. nighttime). Significance was determined at the $p<.05$ level. Table 3 summarizes the results. The effects from schedule flown or flying condition at landing were not statistically significant for any of the variables measured.

Table 2

Mean Scores for Simulated Landing Variables

		M/SD					
Schedule	Fly_cond	DME RMS	GS RMS	LZ RMS	IAF alt	061^0 alt	037^0 alt
A	Daylight	0.41/0.12	0.46/0.12	0.26/0.27	94/131	38/50	50/56
A	Nighttime	0.44/0.11	0.45/0.16	0.22/0.13	129/203	58/106	37/40
B	Daylight	0.46/0.16	0.48/0.11	0.35/0.49	152/250	101/171	46/25
B	Nighttime	0.46/0.12	0.48/0.09	0.27/0.20	161/269	65/88	35/49

Note. Fly_cond = flying condition; DME = Distance Measuring Equipment; GS = glideslope; LZ = localizer; alt = altitude. DME measurements are in nautical miles, GS and LZ measurements are in degrees, and all altitude measurements are in feet.

Table 3

Analysis of Variance for Simulated Landings

		F					
Source	df	DME RMS	GS RMS	LZ RMS	IAF alt	061^0 alt	037^0 alt
				Between	subjects		
Schedule	1	2.01	0.08	0.80	0.72	0.96	.18
Error	14	(0.025)	(0.027)	(0.221)	(73801.64)	(41619.94)	(2141.62)
				Within	subjects		
Fly_cond	1	0.07	0.00	0.89	0.02	0.95	1.94
Schedule x Fly_cond	1	0.48	0.52	0.24	0.08	2.49	0.02
Error	14	(0.018)	(0.011)	(0.125)	(64511.26)	(9804.03)	(1917.44)

Note. Values enclosed in parentheses represent mean square errors. DME = Distance Measuring Equipment; GS = glideslope; LZ = localizer; alt = altitude; Fly_cond = flying condition.

Discussion

Numerous studies have shown that performance on many memory, cognitive, and perceptual-motor tasks are strongly correlated to an individual's circadian rhythms, with the lowest performance occurring during the circadian trough (approximately 0400 hours-0500 hours) (e.g., Kleitman, 1938/1963; Tilley & Brown, 1992; Smith, 1992). Schedule A required the pilots to deploy at 2100 hours and land at the deployed site at 0600 hours. Also, the scheduling of the six subsequent engagement sorties kept the pilots in conflict with their normal sleep/wake cycle, thus increasing cumulative fatigue throughout the week. Conversely, Schedule B was off-set twelve hours from the nighttime schedule, and was considered more circadian "friendly". It required the pilots to deploy at 0900 hours and land at 1800 hours. For these reasons, it was hypothesized that landing performance would be worse for those performing Schedule A. Subjective fatigue ratings taken throughout each week showed the pilots on Schedule A felt more fatigued than those on Schedule B, however the objective landing data did not show any significant fatigue effects.

A possible explanation for these results could be the high experience level of the subjects. Each pilot had received enough training and exposures to actual landings as to mask any performance decrements caused by fatigue. This means that even though the pilots were fatigued, they could land the plane safely and operate within prescribed Air Force operational parameters. This might imply that the measurements used were not sensitive enough to capture any fatigue effects. However, we were interested in the pilots operational flying performance and how this responded to fatigue. Additional follow-on studies could be conducted that look at the issue of fatigue over a longer period of time (i.e., more sorties after deployment).

Although the study did not show any differences between the two schedules, the results can make an argument for modifying one of the Air Force's regulations. Currently, pilots must land at a deployed site during the daylight. For eastward deployments, this means they must take-off and land at times very similar to this study's Schedule A. Polling the subjects showed 82% of the pilots would rather conduct a daytime deployment, which would require a nighttime landing at the deployed site. The results obtained in this study indicate no performance difference between daylight and nighttime landings, and possible consideration could be given to changing the regulation to allow nighttime landings at deployed sites. This would allow for greater flexibility in mission planning, based upon the specific situation surrounding each deployment, and allow pilots to fly a more circadian friendly deployment when operationally sound.

References

Kleitman, N. (1963). Sleep and wakefulness (rev. ed.). Chicago: University of Chicago Press.

Tilley, A., & Brown, S. (1992). In A. P. Smith & D. M. Jones (Eds.). Handbook of human performance: Vol 3. State and trait (pp. 237-259). San Diego: Academic Press.

Smith, A. P. (1992). Time of day and performance. In A. P. Smith and D. M. Jones (Eds.). Handbook of human performance: Vol 3. State and trait (pp. 217-235). San Diego: Academic Press.

Hemispheric Dominance and Flight Performance

Christopher T. Johannssen
Anthony J. Aretz
United States Air Force Academy

Abstract

Existing research suggests the two hemispheres of the brain show significant information processing advantages relative to one another. The left hemisphere shows an advantage in verbal or analytic tasks while the right hemisphere shows an advantage in visual/spatial tasks. The present study investigated the affect of hemispheric dominance on flight performance. The performance of 27 subjects was evaluated after flying a part task flight simulator. The data revealed that right hemisphere dominant subjects performed better than left hemisphere dominant subjects. Establishing hemispheric dominance as a discriminator of flight performance implies that perhaps it could be used as a viable predictor of success in flight training.

The cortex of the brain is anatomically divided into two symmetrical hemispheres. Although these hemispheres are relatively symmetrical in a physical sense, they are not completely equivalent in their cognitive abilities (Hellige, 1993). Evidence of hemispheric asymmetry is most strongly supported by split-brain patients whose right and left hemispheres are no longer able to communicate and allow investigators to define the asymmetries of the left and right hemispheres. These and other data have lead researchers (e.g., Springer and Deutsch, 1993; Hellige, 1993) to suggest that the left hemisphere processes information in a sequential and analytic manner and is predominantly involved in the production and understanding of language. Meanwhile, the right hemisphere processes information holistically and is responsible for visual-spatial skills. The left hemisphere is typically dominant for a number of important aspects of language, including overt speech, phonetic decoding, syntactic and semantic processing. The right hemisphere is dominant for visual-spatial processing and manipulating objects in three dimensional space. Although researchers have described the specific cognitive abilities of the left and right hemispheres in various ways, they do not necessarily agree on how to characterize their differences.

Despite the fact that verbal and visual-spatial resources seem to be hemisphere specific, both hemispheres do have some ability in most tasks. Springer and Deutsch (1993) for example, have demonstrated that the right hemisphere can show comprehension of certain words, especially object nouns. Consequently, the right hemisphere is capable of contributing to verbal processing by using a less efficient visual strategy, such as shape recognition of letters or words.

Since there are significant differences in the task specific abilities of each hemisphere, is it not reasonable for the most capable hemisphere to take the lead in information processing? According to Hellige (1993) the left and right hemispheres do share information processing control depending on hemispheric ability. The term Hemispheric Dominance (HD) is used to refer to the degree each hemisphere tends to assume control of information processing (Hellige, 1993).

One way of measuring HD is by presenting the subject with a task that stimulates each hemisphere. By comparing the response times of each hemisphere, it can be determined which hemisphere dominates the information processing strategy employed by the subject For example, a subject who demonstrates a tendency to accomplish a language task more quickly when it is presented to the right hemisphere suggests that the subject is using a visual strategy to accomplish the task. Such a subject would be considered Right Hemisphere Dominant (RHD). Similarly, those subjects who respond more quickly when the stimulus is presented to the left hemisphere is most likely using a verbal strategy to accomplish the task. Such a subject would be considered Left Hemisphere Dominant (LHD). Equally Dominant (ED) subjects would be expected to fall some where in the middle employing both verbal and visual-spatial strategies. Unlike the LHD or RHD subject, ED subjects would not have a preferred strategy.

Although Springer and Duetsch's (1993) research suggests that the differences in inter-hemisphere performance are quite small (normally on the order of only a few percentage points for identification and only a few milliseconds in response time), these differences can indicate the degree of HD in a subject's brain. Hence, we were interested in what the costs and/or benefits of these differences would be in a complex task. Most existing hemispheric dominance research is based on simple laboratory tasks.

The present study was designed to examine how hemispheric dominance affects flight performance. The flying environment is cognitively rigorous. Flying also requires a pilot to maintain a three dimensional mental model of where they are relative to other aircraft, targets, and ground references. Thus, the flying environment places a significant demand on the visual-spatial resources of the right hemisphere. Knowing the right hemisphere is dominant for visual-spatial tasks, it is reasonable to propose that hemispheric dominance could have a significant affect on flight performance. Hence, the present study investigated the hypothesis that RHD subjects should perform better in flying tasks.

Method

Participants

Twenty-seven USAF Academy cadets (24 males and 3 females, 19 to 22 years old) voluntarily participated in the study. The subjects, selected from an initial pool of forty volunteers, were selected and classified via hemispheric dominance based on their response times to a split visual field Sternberg task. There were nine LHD subjects, nine Equally Dominant (ED) subjects, and nine RHD subjects. The LHD and RHD subjects were selected based on their position in the most extreme tails of the distribution of response times. The nine ED subjects were selected based on their neutral position in the distribution. Eighteen subjects had no flying experience and the other 9 had 2 to 106 hours of prior flying time, with an average of 15.0 hours.

Apparatus

PES. An Israeli Aircraft Industries (IAI) Pilot Evaluation System (PES) was used to present subjects with different flight scenarios containing up to five concurrent tasks (see Figure

1). The PES was specifically designed to assess fighter pilot potential (IAI, 1991). The PES received preliminary validation as a flight screener by Garvin, Acosta, and Murphy (1995). Subjects used the radar display and controls to perform up to five different concurrent tasks (see Figure 2). These tasks included: target intercept, matching target altitude, matching target velocity, pointing at the target, and tone response.

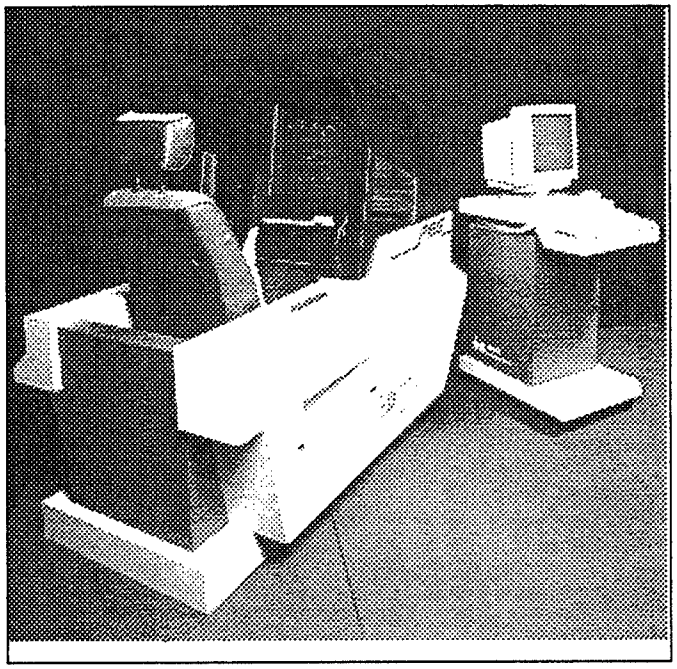


Figure 1. PES hardware.

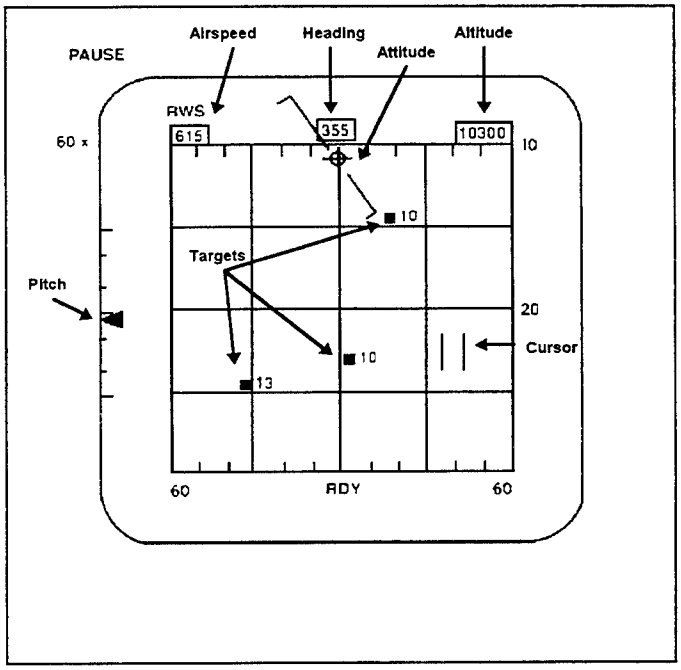


Figure 2. Typical Radar Screen.

Target intercept. Intercepting and launching a missile at a target was considered the primary task.. After the subject locked on a target using a cursor control switch on the throttle, a more detailed view of the target was activated. When the target was within 18 miles, the subject was required to disengage and re-engage the target using two switches. When the target was within the firing envelope (depicted on the radar screen), the subject armed and fired the missile.

Match target altitude. Subjects manipulated the stick to match target altitude. Target altitude was displayed next to the target on the radar display.

Match target velocity. Subjects manipulated the throttle to match and maintain target velocity. While maintaining both the altitude and velocity of the target, both the stick and throttle inputs were coupled to affect altitude and velocity as in a generic fighter aircraft.

Point at target. Subjects used the stick to maneuver the aircraft to match the target's heading so that it remained in the middle of the display.

Tone response. The subject responded to random tones (high or low) presented through a headset at approximately 15-30 sec intervals. Subjects pressed the right button on the throttle for a high tone and the left button for a low tone.

Overall performance. Overall performance on each flight scenario was computed by the PES using a figure of merit technique based on ideal performance (derived from the performance of Israeli fighter pilot baselines, IAI, April, 1995). The principles used in the computation of the figure of merit were: 1) subjects received points for staying within a desired performance window and lost points for deviations; 2) side tasks were weighted more heavily than the primary target intercept task, resulting in a poorer score if the side tasks were not performed well. The final score was a single percent score based on the subject's score divided by the total possible.

Auditory Digit Task

In addition to the PES flight tasks, subjects were also asked to perform an auditory digit task. A tape player was used to present single digits at a rate of one digit every four seconds. Subjects were required to respond vocally with the absolute value of the difference between the digit just presented and the previous digit. Consequently, the task required subjects to keep a single digit in working memory. Digit performance was scored using the percent correct responses during a scenario flight.

Scenarios

Table 1 shows the nine scenarios used in the study. These scenarios were created by combining eight PES scenarios selected from a possible 22. These nine scenarios were created by combining eight PES scenarios (selected from 22 possible PES scenarios) with the digit task for a total of nine scenarios containing from one to six concurrent tasks in different combinations. These nine scenarios were created to include a variety of task combinations.

Table 1. *Scenario Descriptions.*

Scenario	Tasks	Total Tasks
1	digits task	1
2	intercept target	1
3	digits+intercept/match target velocity	2
4	match target altitude/velocity +keep target altitude	3
5	digits+tones+intercept/match target velocity	4
6	tones+intercept/match target altitude/velocity	4
7	digits+tones+intercept/match target altitude/velocity	5
8	tones+intercept/match target altitude/velocity +keep target in center	5
9	digits+tones+intercept/match target altitude/velocity +keep target in center	6

Procedures

Subjects were presented with a computerized tutorial explaining the operation of the PES. Next, subjects flew three practice scenarios in the same sequence for familiarization with the PES

tasks. Subjects then completed nine data collection scenarios. The order of the data collection scenarios was counterbalanced using a modified Latin square.

NASA-TLX ratings were collected at the completion of each scenario (Hart and Staveland, 1988). At the conclusion of all scenarios, subjects completed a paired comparison of the six subscales of the NASA-TLX, generating a weight for each dimension used to compute an overall workload rating. Each subject took approximately one hour to complete the study.

Results

A two-way mixed design ANOVA was performed on the PES performance scores using Hemispheric Dominance (HD) and PES scenario as the independent variables and PES score as the dependent variable. The results showed that HD and the PES scenario had a significant effect on flight performance, $F(2,24)=3.62$, $p=.042$, $F(7,168)=13.53$, $p<.001$ respectively. The interaction was not significant. A post hoc analysis revealed RHD subjects had a significantly higher average PES score ($M=21.3$) than LHD subjects ($M=15.5$) (see Figure 3). Equally dominant subjects had an average score of 19.7. An additional analysis of the TLX ratings revealed no significant differences. There were also no significant differences on the digit task.

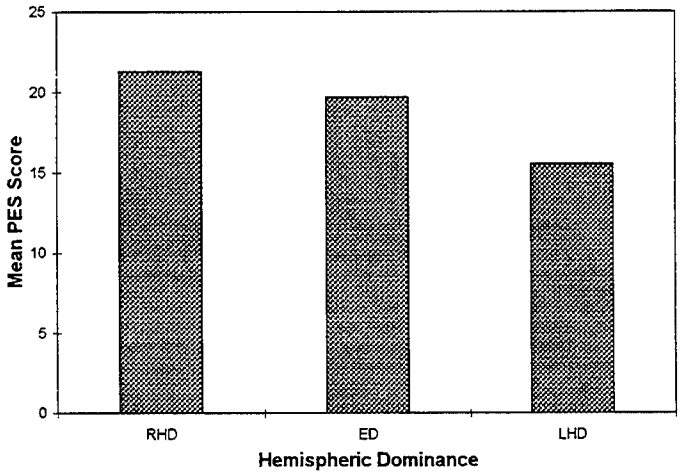


Figure 3. PES Performance by Hemispheric Dominance.

Discussion

The hypothesis of this study, that RHD pilots would display higher flight performance was supported. RHD subjects performed better on the PES than LHD subjects. Springer and Duetsch's contention that inter-hemispheric performance differences are relatively small would lead us to believe that any difference we might expect in flight performance would be negligible (1993). On the contrary, the results indicate these hemispheric differences have emerged as a discriminating factor in flight performance.

The results not only support the hypothesis that RHD should perform better in flying tasks, but it also suggests that the proposed logic of why RHD should perform better is correct. RHD subjects (M= 21.3) performed significantly better than LHD subjects (M=15.5). Since there is no significant difference in performance based on the NASA TLX or prior flight time, the only significant difference between the RHD and LHD subjects is their hemispheric dominance. Given that the LHD subjects have less efficient visual-spatial processing, as compared to RHD subjects, the results suggest that RHD subjects perform better because they have more efficient visual-spatial processing resources. The results also suggest that ED subjects fall somewhere in the middle because ED subjects employ both visual-spatial and verbal strategies. These data are consistent with Friedman and Polson's (1981) hypothesis that the two hemispheres act as independent resource pools for information processing tasks. It is important to note, however, that it cannot be conclusively deduced from these results that the visual-spatial demands of the flying tasks are truly those specific demands that are affected by differences in hemispheric dominance. Hence, the present study represents only the beginning of a series of experiments that would be required to determine exactly why RHD leads to such an advantage in flight performance.

Finally, despite the need to conduct further experimentation to resolve the issues discussed above, the simple fact that hemispheric dominance has been identified as a discriminator for flight performance has significant implications. The results suggest that individual differences in hemispheric dominance may be a viable diagnostic tool for predicting who has the best chance of success in flight training. Using hemispheric dominance in concert with the tools already employed to predict pilot success may serve to allow even more accurate predictions. As it becomes increasingly cost prohibitive to send candidates to flight school who may or may not succeed, a more accurate method of predicting flight performance could prove to be valuable. A longitudinal study tracking undergraduate pilot training students based on their hemispheric dominance would be necessary to determine the diagnostic value of such a measure.

Conclusion

This study conducted research on the affects of hemispheric dominance on flight performance. It was determined that flight performance was affected by hemispheric dominance. Specifically, RHD subjects performed significantly better on simulated flight. This finding suggests that hemispheric dominance has the potential to be a robust predictor of flying performance. Further research is needed to examine the validity of this possibility.

References

Burel, B. & Kahn, M.V.A. (1992). Quality curriculum for the brain. Education, 113, 240-246.

Friedman, A. & Polson, M.C. (1981). Hemispheres as independent resource systems: Limited capacity processing and cerebral specialization. Journal of Experimental Psychology: Human Perception and Performance, 7, 1031-1058.

Garvin, J.D., Acosta, S.C., & Murphy, T.E. (1995). Flight training selection using simulators - A validity assessment Presented at the Eighth International Symposium on Aviation Psychology. Columbus, OH: Ohio State University.

Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.S. Hancock & N. Meshkati (Eds.), Human Mental Workload, pp. 139-183. Amsterdam: Elsevier.

Hellige, J.B. (1993). Hemispheric asymmetry: What's right and what's left. London: Harvard University Press.

Israeli Aircraft Industries. (1991) Pilot Evaluation System Validation and Design Principles. [Brochure]. Lod, Israel: Author Springer, S.P., & Deutsch, G. (1993). Left brain, right brain (4th ed.). New York: W.H. Freeman and Company.

Virtual Reality Features of Frame of Reference and Display Dimensionality with Stereopsis: Their Effects on Scientific Visualization

2Lt Edward P. McCormick
Edwards AFB, CA.
Christopher D. Wickens, Ph.D.
University of Illinois at Urbana-Champaign

Abstract

Dimensionality, stereopsis, and frame of reference are modified in an experiment that contrasted performance using a 2D display, and four displays varying in frame of reference (immersed vs. non-immersed) and in the presence of stereoscopic vision. Performance was measured across two separate scientific visualization subtasks: local and global judgment support. Participants were instructed to locate and follow a designated path through simple virtual environments and to answer questions about that environment. The results revealed that 2D performance was substantially worse than 3D performance across both frames of reference and stereo conditions. The results also indicate that the immersed frame of reference severely hampered global judgment support. Local judgment accuracy benefited from stereo, but response time was unaffected. Global judgments showed no accuracy advantage from stereo, but did show faster responses when stereo was used.

The rapid increase in interest in scientific visualization has led many to search for "the ultimate form" of visualization medium. While some 3D computer displays have enhanced data visualization (Wickens, Merwin, & Lin 1994), others have interpreted "well designed" to necessarily mean advanced technology and most recently Virtual Reality (VR). This has led to the debatable assumption that advanced (rather than properly integrated) technology invariably holds the key to enhanced visualization.

VR is not a unified phenomenon, but rather can be broken down into separate analyzable features. The goal for this research was to examine the effects of certain features of VR; namely Dimensionality (whether a 3D data set should be presented as a 3D volume or 2D co-planar displays), stereoscopic augmentation, and frame of reference (Immersion vs. fixed view exocentrism) on data visualization, specifically the support of correct judgments (local & global) about data sets.

Local judgments can be characterized by how well the displayed information facilitates the user's evaluation of an object's position. This position can also be characterized either as relative (in relation to other objects) or as absolute (i.e., measured along a single scale, such as individually measuring heat, volume, depth, etc.). This experiment focused on only the relative judgment support. Summarizing the dimensionality effects on local judgment support studies suggest that 3D displays are no better (or actually are worse) than 2D displays in local judgment accuracy, but do support faster responding (Wickens & Prevett, 1995). The research on

stereoscopic viewing (Barfield & Rosenburg, 1995) indicates that stereo significantly enhanced the performance of local judgment tasks. Based on this research we still are unclear about which level of dimensionality (2D vs. 3D) will best support local judgments. However, when considering frame of reference (FOR) Wickens and Prevett found that increasing exocentrism supported better local judgments. We expected that performance on local judgments would best be facilitated by the non-immersive 3D view. Furthermore, the stereoscopic viewing should have added further precision in estimating objects' positions in the database.

Global judgments can be thought of as measures of the viewer's ability to meaningfully understand the distribution, shape or size of the environment that she or he just encountered? Research on dimensionality and its effects on global judgments in aviation contexts suggests that 2D displays will support better global judgment performance (Rate & Wickens, 1993). However Wickens, Merwin and Lin (1994) showed that when using displays for data visualization, 3D displays allowed for better integration of information. Furthermore, they found that stereoscopic viewing further improved global judgment performance. In examining how FOR will affect global judgment support, research suggests that exocentric points of view will better support global judgments than will egocentric points of view (Aretz, 1991 and Wickens & Prevett, 1995). In summary, we expect the non-immersive 3D display to have the best performance on global judgment tasks. Furthermore, the stereoscopic viewing should add further distinction of the data bases' true distribution or shape.

Method

Subjects

Thirty students at the University of Illinois (15 males and 15 females), all with normal or corrected to normal vision, served as voluntary participants. All of the subjects were paid $5.00 an hour for their participation. Two groups for the between subjects manipulation of FOR were formed by randomly assigning eight males and seven females to one group and conversely seven males and eight females to the other group.

Apparatus

All subjects were required to execute a navigation task rendered in 3D space on a Silicon Graphics IRIS workstation. There were three display perspectives used to guide the subjects in the 3D space; 2D, 3D immersed and 3D non-immersed. Stereoscopic viewing was produced by the use of Silicon Graphics Stereoview glasses. Initially all subjects were to complete a total of 15 trials composed of 5 trials of each of the three possible perspectives (for later subjects, the 2D condition was deleted for reasons that will be explained below): 2D split screen perspective, a monoscopic perspective and a stereoscopic perspective. The later two perspectives were represented in either an immersed or non-immersed FOR. The actual data environment was represented as a cube volume with six different colored walls. The volume, with 6 uniquely colored walls, contained a series of 15 major objects (destinations) along with 200 randomly placed smaller objects. The 15 major objects were defined within each cube to represent a particular 3D pattern.

Procedure

All subjects were instructed to navigate the 3D space as quickly and with as little joystick input as possible. The actual task required the subject to navigate the target icon. This icon was represented by a small arrow head symbol in one (3D) or both (2D) panels of the non-immersed conditions and by the display viewpoint defining the subject's field of view for the immersed condition. This icon was to be navigated by way of a joystick through the 3D space along a path designated by flashing target cubes. The program would automatically terminate that particular trial after interception of all 15 target cubes. During each navigation trial the program would periodically halt to ask the subject to make a precise relative judgment of the location of an object within the current field of view (as defined for the immersive condition); identical questions were asked for the 2D and 3D non-immersive displays. This would take the form of a multiple choice question asking the subject which of the two objects was closer to his or her own position or to a certain colored wall. Furthermore, after completing each navigation trial the program also asked a single multiple choice question to assess the subjects' knowledge of the global pattern or distribution of the particular data points just encountered. This assessment began by blanking out the entire screen and asking the subject a question regarding the general shape of the data base.

Results

Two separate analysis were performed in this experiment. The first analysis consisted of a one way ANOVA which compared the three frames of reference against one another (2D, immersive and non-immersive), the latter two only within the monoscopic viewing condition. Two Tukey post-hoc tests were conducted on significant FOR main effects. All significant differences are at an alpha level of .05 and all non-significant differences are at an alpha level of .10. The second analysis was a 2X2 design which analyzed the effects of stereoscopic viewing across the immersed and non-immersed frames of reference. The 2D display condition proved to be very difficult in some phases. Preliminary analysis of the results from the seven (out of fifteen) subjects that completed the 2D trials indicated that the performance in the 2D condition was significantly worse. Therefore, the remaining 23 subjects did not perform the 2D trials.

Local Judgment

The average accuracy of local judgment tasks is shown in Figure 1. The FOR analysis showed a main effect $F(2,327) = 28.90$, $p<.0001$, revealing that the 2D displays were the least accurate viewing condition. The FOR Tukey analysis indicated that the 2D condition had significantly lower accuracy than the non-immersed and immersed conditions. There was no significant difference in accuracy between the non-immersed condition and the immersed conditions. A similar analysis of stereo effects revealed that performance in the stereo conditions was significantly more accurate than in the non-stereo conditions $F(1,296) = 3.86$, $p< .05$. Average response times on local judgment questions are shown in Figure 2. The FOR ANOVA indicated a significant main effect $F(2,324) = 12.87$, $p<.0001$, revealing that the 2D displays had by far the slowest response time to the local judgment questions. The FOR post-hoc analysis revealed that the 2D condition was significantly slower than the non-immersed and immersed

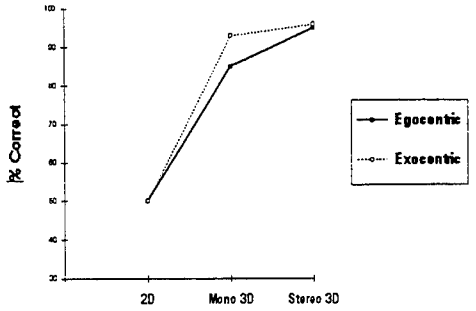


Figure 1. Accuracy of local judgments

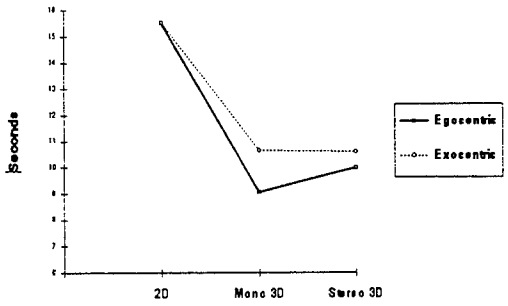


Figure 2. Local judgment response time

conditions. There was no significant difference in response time between the non-immersed and immersed FORs. The stereoscopic ANOVA revealed no significant difference between the two viewing conditions ($F = .44$, $p = .50$).

Global Judgment Support

Accuracy on the global judgment tasks is shown in Figure 3. The FOR analysis revealed a significant main effect $F(2,327) = 34.49$, $p<.0001$, indicating that the non-immersed displays had the highest global judgment accuracy. The FOR analysis indicated that the non-immersed condition had a significantly higher accuracy than did the 2D and immersed conditions. There was no significant difference in accuracy between the 2D displays and the immersed displays. The stereoscopic ANOVA revealed no significant difference in global judgment accuracy ($F = .29$, $p = .59$). Average response time on global judgment questions are shown in Figure 4. The FOR ANOVA revealed no significant main effect ($F = .97$, $p = .38$). The stereoscopic ANOVA revealed stereoscopic viewing to promote significantly faster response times than did non-stereoscopic viewing $F(1,285) = 5.45$, $p<.02$.

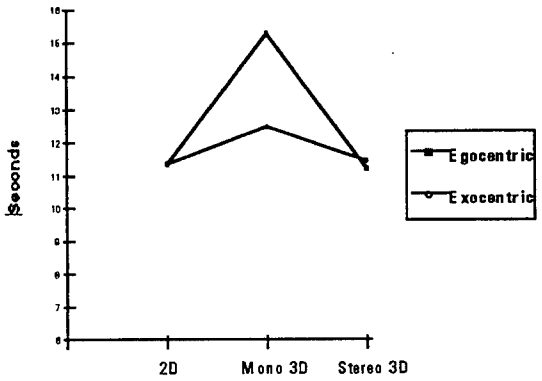


Figure 3. Global judgment accuracy

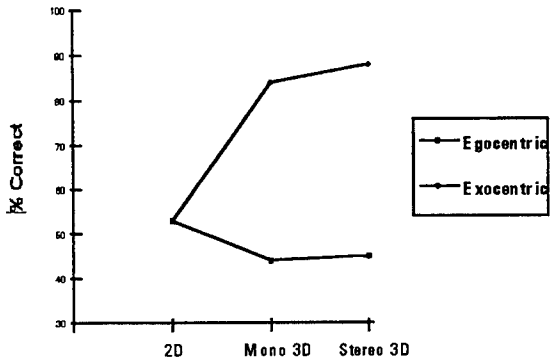


Figure 4. Global judgment response time

Discussion

Local judgment support was severely hampered by 2D viewing, which produced slower response times with less accurate responses. Global judgment support data indicated that 2D accuracy was substantially worse than the non-immersive 3D view, while 2D viewing did produced faster responses than 3D viewing. These results suggests any task requiring integration will be degraded by separated 2D display. The non-immersed (exocentric) FOR supported superior global judgments. Stereoscopic viewing failed to improve global judgment support, but improved local judgments. It is understandable that stereoscopic viewing should not enhance memory with an immersed FOR, because even a "key hole" view that provides improved depth perception (via stereo) is still a key hole view. In terms of global judgment support, it is more likely that the small increments in depth judgment accuracy, fostered by the stereo viewing, were of insignificant magnitude to substantially add to global understanding.

References

Aretz, A. J. (1991). The design of electronic map displays. Human Factors, 33(1), 85-101.

Barfield, W., & Rosenburg, C. (1995). Judgments of azimuth and elevation as a function of monoscopic and binocular depth cues using a perspective display. Human Factors, 37(1), 1-9.

Rate, C., & Wickens, C. D. (1993). Map dimensionality and frame of reference for terminal area navigation displays: Where do we go from here? University of Illinois Institute of Aviation Technical Report (ARL-93-5/NASA-93-1). Savoy, IL: Aviation Research Lab.

Wickens, C. D., Merwin, D. H., & Lin, E. L. (1994). Implications of graphic enhancements of scientific data: Dimensional integrality, stereopsis, motion and mesh. Human Factors, 36(1) 44-61.

Wickens, C. D., & Prevett T. T. (1995). Exploring the dimensions of egocentricity in aircraft navigation displays. Journal of Experimental Psychology: Applied, 1(2), 110-135.

Spatial Knowledge Acquisition in Virtual Environments

Michael J. Singer, Ph.D.[1]
Robert C. Allen
U.S. Army Research Institute for the
Behavioral and Social Sciences

Abstract

Virtual Environment (VE) technology provides a new way to simulate real world activities. This technology will enable the Army to plan, train, and rehearse both individual and collective dismounted soldier tasks. Spatial knowledge of the terrain is a fundamental requirement for many of these tasks and activities. We conducted an experiment in which three groups of subjects "moved" through simulated terrain, learning landmarks: a High-level Virtual Environment (Hi-VE) group, a low-level group (Lo-VE), and a control group (using topographical maps). Results indicate that acquisition activities in Hi-VE produce significantly better spatial knowledge than the same activities using topographical maps.

Virtual Environment (VE) technology provides a new way to simulate real world activities, which will enable the Army to simulate activities for dismounted infantry. The available technology is not currently sufficient for exact real world replication, nor for allowing the replication of the entire range of normal soldier interactions with the real world (Jacobs, Crooks, Crooks, Colburn, Fraser, Gorman, Madden, Furness, and Tice, 1993). The Army Research Institute for the Behavioral and Social Sciences, Simulator Systems Research Unit at Orlando, Florida, has an ongoing research program on the use of Virtual Environments (VE) in training. The focus of the program is the investigation of VE-based learning and transfer effectiveness of both individual and collective dismounted soldier tasks. The primary interest is in small-group leader tasks, subtasks, and activities (e.g., platoon leaders). These tasks have a common context of individual combatants who need to move, observe, shoot, and communicate. The early phases of our research program have been investigating how these four basic activities are learned in VE.

A review of soldier Army Training and Evaluation Program (ARTEP) tasks identified major activities that could be performed, trained, or practiced in VE (Jacobs, et al., 1993). Many terrain interaction activities (e.g., identify safe and danger areas) had high combined rankings in good cost/transfer effectiveness, current technological capability, and commonality of activities across ARTEPS. Underlying many of these activities is the interaction of terrain appreciation skills and spatial knowledge of the operational terrain possessed by the soldier. Terrain appreciation is having a general understanding of how to use terrain characteristics in performing soldier tasks such as weapon's emplacement, defensive positions, and land navigation. Obviously, the effective application of terrain appreciation is dependent on the level of spatial knowledge of the soldier.

Researchers have investigated the representational structure of spatial knowledge and conditions of acquisition (Siegal & White, 1975; Goldin & Thorndyke, 1981). As with many

knowledge structures, memory for the environment seems to be built upon increasing levels of elaboration. There are three general knowledge levels that have been described as a result of spatial memory research (Goldin & Thorndyke, 1981). The base level is knowledge about landmarks, the intermediate level consists of knowledge about routes, and the highest level of organization is survey or configuration knowledge that relates sets of landmarks and routes by direction and distance (Goldin & Thorndyke, 1981; Siegal & White, 1975; Witmer, Bailey, & Knerr, 1995).

This experiment investigated the potential improvement in spatial knowledge acquisition in VE configurations over comparable map study, and the possible difference between two levels of VE interface configurations. The lo-VE configuration presented stereographic views in a head-mounted display (HMD) with gaze control and movement controlled by joystick. The lo-VE is akin to fixed-view, joystick directed video-disk and computer-based training, which has been shown to be ineffective for some spatial orientation activities (Lickteig & Burnside, 1986). The Hi-VE configuration linked view to head movements, and controlled movement by walking on a treadmill. The Hi-VE is an improvement on the normal VE head-tracked, joystick-controlled configuration, which has shown transfer of VE-acquired spatial knowledge to real building interiors (Witmer, Bailey, & Knerr, 1995). The expectation is that increasing the psychophysical functionality (more normal gaze and movement control) should lead to a more accurate or complete spatial representation of simulated terrains than is found with Lo-VE simulation or topographical maps.

Method

Eighteen females and thirty males were recruited from the University of Central Florida, for a total of 48 subjects. Subjects ranged in age from 18 to 44 with a mean age of 24.6. Subjects were required to pass a battery of standard vision tests (corrective lenses were allowed) before participation. Subjects passing the eye exam were given introductory training on topographical maps, terrain features, and threat identification. Subjects were then tested on their knowledge, and those not meeting minimum requirements were excluded from the experiment. All subjects were paid $5.00 per hour or given course credit for their participation.

This experiment was performed at the University of Central Florida, Institute for Simulation and Training (IST). The visual display information was generated using Performer^{tm} and adjunct specialized software developed by IST, by a Silicon Graphics ONYX^{tm}. The visuals were presented through a Virtual Research Systems VR4 Head Mounted Display (HMD). The VR4 has 48°x36^{o} field of view, with 742x230 color pixels in each lens (Real Time Graphics, 1995). Head and hand pointing were tracked by Polhemus Isotrak^{tm} sensors. The treadmill was instrumented for the Hi-VE condition, and provided a constant walking pace. The Lo-VE system used the same HMD without head-tracking and controlled movement by a joystick (movement was set to the same constant walking pace as the treadmill).

The three groups of subjects "moved" through two simulated terrains, performing simple cognitive terrain appreciation activities. The Hi-VE walked on the treadmill using a head-tracked HMD and pointing to indicate direction selection. The Lo-VE group moved through the same

simulated terrain, performing the same activities; while seated and observing the terrain through a non-head-tracked HMD, using a joystick for movement and the pointing hand for direction indication. The control group (Map) performed the activities using topographical maps, with paced study replacing the movement through terrain, while seated at a desk. After the practice session, during which subjects followed a designated route and learned the landmarks, subjects' configuration knowledge of the terrain was tested in the same condition in which they practiced, and the Map condition transferred to the Hi-VE configuration. The test placed the subjects at previously unvisited sites and had the subjects point at requested landmarks. Some of the landmarks were visible and some not visible from the sites, with visible and non-visible landmarks varying by site.

Results

Each landmark indication was scored as correct if the directional indication was within the angle subtended by the visible feature from the tested position, or for very narrow landmarks within +/-2.5^{O}. The range allowed for non-visible landmarks was +/- 22° of the center point for the landmark. The number of correctly identified landmarks was then summed for each subject at each test site.

A repeated measures ANOVA using the number of correctly identified landmarks found significant differences over the experimental conditions ($F=4.27$, $p=.021$; Hi-VE=3.27, Lo-VE=2.81, Map=2.26). A Post Hoc analysis of the experimental conditions found only the difference between the Hi-VE condition and the Map condition significant (HSD; Difference=1.01, $p<.05$). The mean number of correctly identified landmarks also differed significantly over the six sites ($F=14.81$, $p<.001$), but was not significantly different over terrains. No interactions of conditions, terrains, or test sites were significant.

An ANOVA of only visual landmarks over all test sites did not show any difference between terrains. There was a significant difference between experimental groups ($F=4.83306$, $p=.013$; Hi-VE=19.6248, Lo-VE=16.875, Map=13.5624). A Post Hoc analysis found a significant difference (6.0624) between the Hi-VE and Map condition means ($HSD_{.05,\ 16}=3.6673$). An ANOVA of only the non-visual landmarks did not find any significant differences for terrains or experimental groups. There was a significant correlation between correctly identified visible landmarks and correctly identified non-visible landmarks (Pearsons's $r=.6178$, $p<.001$).

Discussion

The central theme of this experiment is the development of spatial knowledge in response to different VE configurations and relative to map exercises. The results indicate that a better knowledge of landmarks was acquired in the Hi-VE with overall landmark identification and identification of only visible landmarks. Better identification of visible landmarks in the VE shows the superiority of visual experience-based spatial learning over the spatial learning from symbolic representations in the Map condition. The significant correlation between the correctly identified visible and non-visible landmarks supports the hypothesis of better spatial knowledge being gained from visually-based acquisition.

The results indicate a better memorial representation of spatial relationships between landmarks, as reflected in more correct identifications of non-visible landmarks given the visible landmarks as cues. This indicates that landmarks are not learned individually, but that very early representations are formed that include angles and distances between landmarks. Humans learn from both symbolic information and experience, and learn some things, such as spatial organizations, better through experience than through symbology (Goldin & Thorndyke, 1981). It is clear that a VE configuration that allows more normal physical (stereoscopic visual displays) and functional (head-slaved visual displays and walking-based movement) interactions promotes better spatial knowledge acquisition.

The findings reported here contribute to our understanding of how soldier's memory for spatial organization is affected by VE experiences. The issue relates directly to activities performed in standard Infantry activities, and is important in developing dismounted soldier simulations for realistic training.

References

(1995). Real Time Graphics. Computer Graphics System Development Corporation, Vol.4(2), Mountain View, CA.

Goldin, S. E., & Thorndyke, P. W. (July 1981). Spatial Learning and Reasoning Skill. Rand Corp. (R-2805-ARMY Contract MDA-903-79-C-0549), Santa Monica, CA, prepared for U. S. Army Research Institute for the Behavioral & Social Sciences, Alexandria, VA.

Jacobs, R. S., Crooks, W. H., Crooks, J. R., Colburn, E., Fraser, II, R. E., Gorman, P. F., Madden, J. L., Furness, III, T. A., & Tice, S. E., (in press). Behavioral Requirements for Training and Rehearsal in Virtual Environments. (TR 1011, AD A286 311), Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA.

Lickteig, C. W., & Burnside, B. L., (1986, November). Land Navigation Skills Training: An Evaluation of Computer and Videodisc-Based Courseware. (Technical Report 729), U. S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA.

Siegal, A. W., & White, S. H. (1975). The development of spatial representations of large-scale environments. In Reese, H. (Ed.), Advances in Child Development and Behavior, Vol. 10, New York, Academic Press, 10-55.

Witmer, B. G., Bailey, J. H., & Knerr, B. W. (1995). Training Dismounted Soldiers in Virtual Environments: Route Learning and Transfer. (Technical Report 1022, ADA ###), U. S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA.

Deriving Training Lessons Learned From an Advanced Warfighting Experiment

Gary S. Elliott, M.A.
U.S. Army Research Institute

Abstract

The U.S. Army is conducting Advanced Warfighting Experiments (AWEs) to determine the force design, equipment, doctrine, and training requirements for Force XXI, the 21st century Army. The AWE Focused Dispatch was conducted to examine the impact of integrating digital systems on a battalion task force (TF) organization, doctrine, and warfighting capabilities. Effective soldier and unit training for emerging digital technologies is essential to the successful conduct of future AWEs and Force XXI training. This research effort documented the TF digital training efforts, captured the lessons learned, and identified implications for future Force XXI training efforts.

The U.S. Army is conducting a campaign to develop the Army of the 21st century - Force XXI. The Army has recognized it must exploit the enhanced capabilities of modern information systems to ensure success in future warfighting operations. To support this effort the Army has instituted a series of AWEs to address hypotheses about doctrine, training, leader development, operations, material, and soldiers (DTLOMS). These experiments are structured to involve soldiers and units while using tactical scenarios in live, virtual, and constructive simulation environments. AWEs are intended to build on lessons learned from previously conducted AWEs.

The AWE Focused Dispatch (FD) was initiated to examine the impact of horizontal digitization on a battalion TF organization, it's processes, and it's doctrine and tactics, techniques, and procedures (TTPs). This AWE consisted of a series of focused sub-experiments conducted in live, virtual, and constructive simulation environments to examine and refine organization, doctrine, and TTP changes that optimize digital systems, information interconnectivity, and communications for warfighting operations.

A necessary prerequisite condition for conducting a successful AWE is training. This training includes proficiency in fundamentals as well as new digital operations. Documentation of the previous mounted force AWE suggests training was problematic for the digitally-equipped battalion TF (U.S. Army Armor Center, 1994). For this AWE, it was deemed crucial to capture and document the training process and any insights appropriate for future Force XXI training efforts. Thus, the focus of this research effort (Elliott, Sanders, & Quinkert, in press) was to (a) document the TF training preparations including digital training efforts, (b) derive training lessons learned, and (c) examine the implications for future Force XXI training efforts.

Method

Participants and Sample

The Battalion Task Force 2-33 Armor, 16th Cavalry Regiment, U.S. Army Armor School was the AWE FD unit. Questionnaire and interview data were collected from a battalion leadership sample consisting of the battalion commander, executive officer, operations officer and assistant, intelligence officer and assistant, personnel officer, logistics officer, chemical officer, five company commanders, scout platoon leader, medical platoon leader, and a maintenance support platoon leader

Digital Equipment

The TF used a variety of digital technologies mounted on different tactical platforms. The focus of the AWE FD and observations centered on the battalion leadership's use of four command and control (C2) systems to plan, prepare, and execute missions: Intervehicular Information System (IVIS), Brigade and Below Command and Control system (B2C2), All Source Analysis System (ASAS), and Initial Fire Support Automated System (IFSAS). The B2C2 and IVIS were the TF's primary C2 digital systems used to collect, manipulate, and disseminate tactical information. Both were used to track the battle plan, keep the commander advised of the current ground situation, and monitor the close battle. The B2C2 also was used to communicate with the brigade cell. The ASAS was a computer-assisted intelligence and electronic warfare processing, analysis, reporting, and technical control system used by TF intelligence personnel to monitor the enemy situation for the task force, advise the battalion commander and executive officer of the enemy situation, and direct the intelligence collection efforts. The IFSAS was an artillery C2 system used to support the TF's indirect fire planning and execution. Interconnectivity between systems was limited to some immediate call for fire relays between IVIS and IFSAS. Interconnectivity between the different digital systems consisted of manually inputting the data from one system into another system.

Training Environments

The TF used a variety of training environments to conduct individual through collective training. A TF digital learning center (DLC) was equipped with six personal computer workstations with IVIS emulation and tutorial software for conducting initial and sustainment training for individuals and teams. IVIS, B2C2, IFSAS, and ASAS were able to be installed in the DLC when needed for sustainment and collective training efforts. The Mounted Warfighting Simulation Training Center (MWSTC), a simulation network facility at Fort Knox, was used by the TF to conduct command, control, and tactical maneuver of conventionally equipped platoons, company team, and TF levels in a virtual battlefield environment. The Mounted Warfare Test Bed (MWTB), a distributed interactive simulation facility, was used to conduct TF missions in a virtual battlefield environment with some digitally equipped force elements. The Janus constructive simulation facility at Fort Knox was used primarily to train leaders, commanders, and staff in combat operations first as a conventionally equipped force and later as a partially digitally-

equipped force during an experimental event. Local training areas at Fort Knox were used to conduct live simulation or field training for platoon through battalion-level training exercises.

Training

The TF training strategy was planned as a "crawl-walk-run" approach with the integration of live, virtual, and constructive simulation training events. Based on a lesson learned from the previous AWE, the TF deliberately planned to train to proficiency in combat fundamentals as well as in digital equipment training. Task Force training started in early January 1995 and culminated in the last AWE FD sub-experiment in August of that year.

The "crawl" phase of training started with formal new equipment training for one company possessing M1A2 tanks (which contained IVIS). A classroom briefing on digital TTPs was given to most members of the TF in early February. Selected members of the TF attended IVIS-only training conducted in their local motor pool and intelligence and operations staff slices attended IVIS workstation training in the DLC. The TF commander, leaders, and staff attended a Janus command post exercise in mid-February to train battle operations as a conventionally equipped TF. Soon afterwards, the TF DLC started training operations through June, allowing each company one day per week to conduct digital operation sustainment training with its resources. In mid-March, most TF leaders, staff, and support personnel received initial classroom hands-on instruction on the B2C2 equipment. The TF participated in structured virtual simulation training as conventionally equipped TF in the MWSTC at the beginning of April. Immediately afterwards the TF and supporting elements participated as a partially digitally-equipped force conducting offensive and defensive missions during a virtual simulation sub-experiment in the MWTB. In late April, one company, selected TF elements, and battalion staff conducted a limited field training exercise with digital C2 equipment. Immediately afterwards the staff, company commanders, and some support elements conducted a digital communication exercise to develop and refine communication procedures.

During the "walk" phase of the training, the TF commander, staff, company commanders, and selected TF elements participated in a JANUS constructive simulation sub-experiment to develop leader and staff operations as a digitally-equipped TF. The last home station training event consisted of a week long TF field training exercise during mid-June. The digitally-equipped TF conducted force-on-force engagements using laser engagement systems against a company-sized conventionally-equipped opposing force during offensive and defensive missions. The planned "run" phase was actual participation in the final culminating sub-experiment referred to as the Live-Virtual experiment.

Procedure

Training information collection for the previous AWE occurred after TF home station training preparation. Although this approach captured useful training insights, it provided limited details that could have provided more focus for follow-on Force XXI training efforts. The approach used for this research and data collection effort was to observe the training process, and

develop questionnaires and interviews aimed at soliciting specific information geared to perceived training benefits and limitations.

The research team's role in the AWE FD was limited to observing home station training and experiment events in live, virtual, and simulation environments. Except for key events, such as new equipment training and initial digital equipment instruction, company-level and below events were excluded from observation. Generally, collection of observation data was loosely organized around guidelines and criteria used in training program evaluations (Kristiansen & Witmer, 1981; Witmer, 1981). The criteria were useful for organizing training observation notes and questionnaire instruments. Training information was collected for training equipment and materials, training environment, training process, and training evaluation.

After all home station training events had been conducted, a training questionnaire was constructed that was keyed to the particular training events and geared toward confirming observations and training lessons learned. The questions were constructed using training evaluation criteria (Witmer, 1981) and from training principles and criteria used in FM 25-101, Battle Focused Training (U.S. Department of the Army, 1990). Questionnaires were administered to selected TF leaders who were primary operators of digital C2 equipment. Questionnaire administration occurred several weeks after the TF returned from the last sub-experiment.

Interview questions were developed to collect participant perceptions about unobserved training, prerequisite digital skills and knowledge, sustainment training, training delivery methods, training distracters, lessons learned, and recommendations. Interviews were conducted after questionnaires had been completed and returned. Interview sessions were taped for later review.

Results

Training lessons learned and implications were classified into nine categories: training strategy, training management, training methods, prerequisite skills and knowledge, digital learning centers, simulation training, training literature, training assessment, and training support. The training lessons learned are too numerous to present in this paper but several key findings can be presented. Key findings include: (a) units should first train to proficiency on combat fundamentals and then train to digital proficiency before integrating into warfighting operations; (b) identify new tasks resulting from digitization and ensure that the tasks are incorporated into training; (c) training technologies and programs need to be explained to unit personnel when introducing them into unit training programs; (d) the level of digital knowledge and skill is dependent on the digital system, it's interface, and the operator's entry level position in the unit; (e) a digital learning center is a key training environment for executing unit digital training and sustainment training; (f) simulation training can be significantly enhanced when structured training programs are applied; and (g) automation officers and support personnel are needed to support digital and network operations and digital training at the battalion level.

Discussion

Training information collection to derive lessons learned from AWEs can be performed after units have conducted training. However, the number, kind, and specificity of lessons learned will be limited when the data collectors and analysts are unfamiliar with the specifics of the unit's training process and events. Using structured guidelines that adhere to training evaluation principles as a method to collect training information during direct observation of training events yields more detailed information. The training information can then be verified by participants in later data collection settings with instruments and interviews tailored to specific events with recall cues added to assist memory.

Training preparation for a unit to participate in any experiment is vital to ensure validity of results. This is especially true for AWEs which introduce new technologies into organizations for evaluating the effects for future force design and warfighting. It is paramount that soldiers and units are trained to operate the digital equipment and be able to leverage the equipment to provide the best opportunity for examining Force XXI issues. Given the critical role that training has in the march toward Force XXI, it is recommended that unit training for AWEs be thoroughly captured during their training and verified after AWE completion to assess the impact on AWE results. Further, lessons learned and insights need to be documented to have an impact on the next AWE and future Force XXI training efforts.

References

Elliott, G. S., Sanders, W. R., & Quinkert, K. A. (in press). Training in a digitized battalion task force: Lessons learned and implications for future training (ARI Research Report). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Kristiansen, D. M. & Witmer, R. G. (1981). Guidelines for conducting a training program evaluation (TPE) (ARI Research Product 81-18). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Witmer, R. G. (1981). A job aid for the structured observation of training (ARI Research Product 81-16). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

U.S. Army Armor Center (1994). Advanced warfighting experiment: Operation desert hammer VI. Fort Knox, KY: Author.

U.S. Department of the Army (1990). FM 25-101: Battle focused training. Fort Leavenworth, KS: Author.

A Strategy for Efficient Device-Based Tank Gunnery Training in the Army National Guard [1]

Joseph D. Hagman
U.S. Army Research Institute for the Behavioral and Social Sciences

Abstract

A strategy is proposed for minimizing the device-based training time required to prepare armor crews of the Army National Guard (ARNG) for on-tank training and live-fire gunnery qualification. Using the Conduct-of-Fire Trainer (COFT) and Abrams Full-Crew Interactive Simulation Trainer (AFIST), efficiency is achieved by training only gunnery engagements subjected to live-fire evaluation, focusing on those engagements not performed to standard, and allocating training time to crews that need it most, as determined through pretesting.

To maximize the payoff from the limited time available for training (i.e., 12 Inactive Duty Training [IDT] weekends and 2 weeks of Annual Training [AT] per year), ARNG armor units plan to shift more of the emphasis of tank gunnery training from a tank-based to a device-based approach (U.S. Army Armor School, 1990). For this shift to be successful, an efficient strategy is needed to help unit trainers determine which device(s) to use, which training and evaluation exercises to conduct, and which proficiency standards to apply. The present research was conducted to provide this strategy.

Approach

Strategy development involved the identification of live-fire gunnery evaluation requirements, determination of device capabilities to support these requirements, and selection of a training and evaluation approach to support efficient device-based skill acquisition.

Live-Fire Evaluation Requirements

Each year ARNG tank crews attempt to qualify on Table VIII, a live-fire exercise used to assess intermediate crew-level gunnery proficiency (Department of the Army, 1993). Table VIII contains the 12 engagements shown in Table 1. Each crew fires 10 of the 12 engagements and must score 700 or more out of 1,000 points to qualify. Because of its importance to crew-level gunnery evaluation, Table VIII qualification was adopted as the goal of strategy development.

Device Capabilities

The COFT and AFIST were selected for strategy inclusion. Both allow crews to use realistic tank controls in response to computer-generated images displayed through tank optics. The COFT supports the training of tank commander and gunner pairs within a simulated crew compartment, whereas the AFIST is appended to a stationary tank to support full-crew training.

[1] This paper is not an official Department of the Army document in its present form.
The software-driven training exercises on each device permit engagement of single and multiple targets under fully operational (i.e., precision gunnery) and degraded mode (e.g., inoperative laser range finder) firing conditions. Table 2 shows which device exercises simulate the targeting conditions of Table VIII engagements.

Table 1
Table VIII Engagements

Engagement	Description
	Table VIIIA (Day)
A1	On defense, engage moving and stationary tank with main gun using the gunner's auxiliary sight (GAS) and battlesight gunnery.
A2	On defense, simultaneously engage stationary BMP (tracked armored personnel carrier [APC]) with main gun and stationary BTR (wheeled APC) with tank commander's (TC's) Caliber .50 machine gun.
A3	On offense, engage two sets of troops with coaxial machine gun using precision gunnery.
A4	On offense and under nuclear, biological, and chemical (NBC) protection status, engage two stationary tanks with main gun using precision gunnery.
A5A	On offense, engage stationary and moving tank with main gun using precision gunnery.
A5S	On offense, engage two moving tanks with main gun using precision gunnery.
	Table VIIIB (Night)
B1S	On defense, engage stationary tank with main gun from a three-man crew configuration using precision gunnery.
B2	On defense, engage two stationary BMPs with main gun using precision gunnery.
B3	On offense and under NBC protection status, engage stationary BMP with main gun and stationary rocket-propelled grenade launcher (RPG) team with coaxial machine gun using precision gunnery.
B4	On offense, engage stationary and moving tank with main gun using precision gunnery.
B5	On defense, engage stationary tank with main gun using GAS battlesight gunnery under external illumination.
B5A	On defense, engage moving tank with main gun using precision gunnery.

The Proposed Strategy

Given identification of the specific engagements fired on Table VIII and the capability of COFT and AFIST to simulate these engagements, the following strategy is proposed to guide tank crew training and evaluation of Table VIII-related engagements on each device.

As shown in Figure 1, the strategy begins with a pretest on COFT to assess tank crew proficiency on simulated Table VIII engagements and to identify those engagements not performed to standard. The COFT-based pretest selected for strategy adoption was developed by Hagman and Smith (in press) to predict the probability of first-run Table VIII qualification.

Table 2
COFT and AFIST Training Exercises Corresponding to Table VIII Engagements

Table VIII Exercises	COFT Training Exercises	AFIST Training Engagements
A1	113, 117	6VIA1
A2	101, 111	---
A3	102, 106	6VIA2
A4	102, 106, 110	6VIA3
A5S	102, 106, 110	6VIA4
A5A	102, 106, 110	6VIA5
B1S	103, 107, 119	6VIB1
B2	105	6VIB2
B3	110	6VIB3
B4	102, 106, 110	6VIB4
B5	113, 117	6VIA1
B5A	105	6VIB5

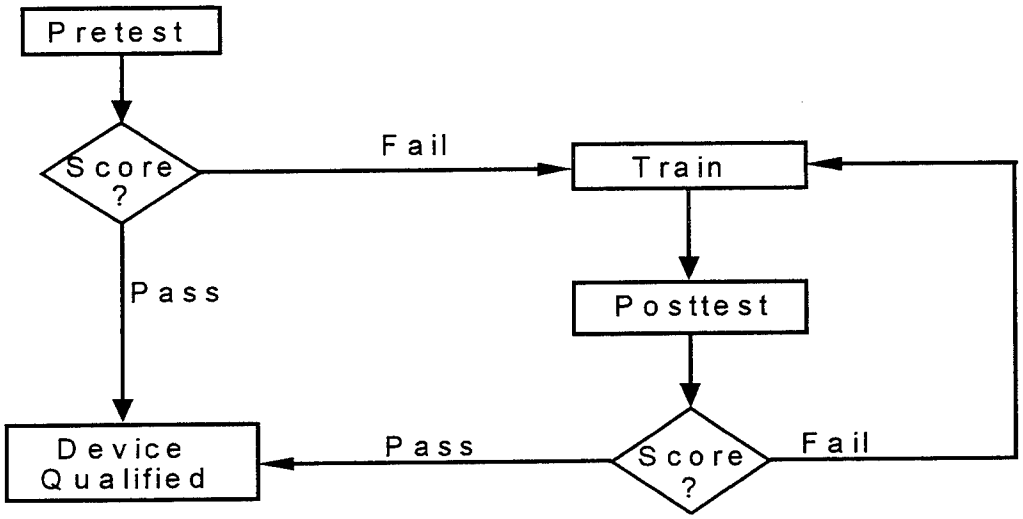


Figure 1. Flowchart depiction of device-based tank gunnery training strategy.

Table 3 depicts a selected range of potential COFT pretest scores in column 1 along with each pretest score's predicted mean Table VIII score and associated probability of first-run Table VIII qualification in columns 2 and 3, respectively. Use of this table enables a unit commander to predict that a particular crew obtaining a COFT pretest score of 765, for example, will on the average fire 700 on Table VIII and have a 50% chance of actual first-run qualification. Based on whether pretest scores are above or below the probability criterion value (e.g., 80%) selected by the unit commander from column 3, some crews will be judged device qualified, whereas others will be judged device unqualified.

Table 3
Predicted Tank Crew Table VIII Score and Probability of First-Run Qualification for Selected COFT Pretest Scores

Mean COFT Pretest Score	Predicted Table VIII Score	Probability of Scoring ≥ 700 on Table VIII
620	562	10%
669	609	20%
706	644	30%
737	673	40%
765	700	50%
793	727	60%
824	756	70%
861	791	80%
910	838	90%

Note. From "Device-Based Prediction of Tank Gunnery Performance," by J. D. Hagman and M. D. Smith, in press, Military Psychology. Copyright by Lawrence Erlbaum Associates, Publishers. Reprinted with permission.

As shown in Figure 1, only device-unqualified crews receive device-based training under the proposed strategy. Thus, training time is devoted only to those crews lacking in gunnery proficiency, thereby promoting efficient allocation of the time available. To promote further efficiency, training is restricted only to specific engagements not performed to pretest standard.

Unlike pretesting, training can be conducted on either COFT or AFIST. As shown in Table 2, COFT can be used to train all Table VIII engagements, whereas AFIST can be used to train all but A2 because the device does not simulate the Caliber .50 machine gun. When both devices are capable of supporting the training of a particular Table VIII engagement (e.g., B1), AFIST should be used as the device of choice because of its ability to promote full-crew integration. When more than one exercise is identified for training a certain Table VIII engagement (e.g., A3), exercises should be alternated to enhance variety and promote transfer.

It is recommended that the provisional standard for crew proficiency on training exercises be set at two consecutive criterion performances. On the COFT, criterion performance is achieved upon crew receipt of an "advance" recommendation from the device in the areas of target acquisition, reticle aim, and system management, as provided on the device's performance analysis printout. On the AFIST, criterion performance is achieved upon crew receipt of a "pass" recommendation from the device for the exercise being trained.

As a final step, crews that have completed training must be posttested (i.e., on the pretest) to ensure that device-based proficiency has been achieved. Crews passing the posttest are considered device qualified, whereas those failing the posttest must return for further device-based training as outlined above.

Implementation Considerations

The proposed strategy is designed for unit implementation over three IDT periods. To promote efficiency, pretesting should be conducted during IDT in conjunction with administration of the Tank Crew Gunnery Skills Test used to certify crew member proficiency on basic gunnery tasks (e.g., identify armored vehicles, load main gun ammunition, issue fire commands).

Before the next IDT period, pretest performance should be reviewed to identify device-unqualified crews and select the appropriate Table VIII-related engagement(s) for training (i.e., those not performed to standard on the pretest). Similarly, the training results of this and the following two IDT periods should be reviewed to select the appropriate exercises for training crews yet to qualify for posttesting and to posttest those that have successfully completed training. Once all crews have passed the pre- or posttest, on-tank training should begin to ensure that crews experience the different aspects of gunnery not practiced or simulated on devices, yet important for successful Table VIII qualification (e.g., open-hatch target acquisition; tank movement and gun recoil effects).

Conclusion

The proposed strategy minimizes the device-based training time required to prepare ARNG armor crews for tank gunnery qualification on Table VIII. Time is saved by restricting training to engagements evaluated on Table VIII, and then only to those not fired to the pretest standard set by the unit commander. Also, by excusing device-qualified crews, training time can be spent on crews that need it most. Posttesting then ensures that previously device-unqualified crews have attained the proficiency level needed for successful transition to on-tank training and live-fire gunnery. Using this strategy, ARNG armor unit trainers can identify which crews to train, which devices to use, which engagements to present, and which proficiency standards to apply for achieving maximum payoff from the limited time available for device-based tank gunnery training.

References

Department of the Army (1993). Tank Gunnery Training (Abrams) (FM 17-12-1-2). Washington, DC: Author.

Hagman, J. D., & Smith, M. D. (in press). Device-Based prediction of tank gunnery performance. Military Psychology.

U.S. Army Armor School. (1990). Armor Training Strategy (ST 17-12-7). Fort Knox, KY: Author.

Training on Simulators and Live Fire Platoon Gunnery Performance

Bruce Sterling, Ph.D.
U.S. Army Research Institute

Abstract

Reduced training resources require the military to increasingly depend on simulators for routine training. Regardless of how inexpensive a simulator may be however, the simulator is useless if it does not enhance performance on the actual equipment. This study demonstrates a relationship between training on platoon gunnery simulators and live fire gunnery performance for US Army tank and Bradley Fighting Vehicle (BFV) platoons. Because these data replicate previous findings for both simulators, results suggest that both tank and BFV platoons may profit from training on platoon gunnery simulators.

Military training will increasingly involve use of simulators and simulations. Due to the worldwide reduction in defense spending, the military will have less ammunition and fewer other resources (e.g., fuel, spare parts, land, hours) for live fire training. Thus, if forces are to maintain the same training tempo and level of combat readiness, the military must use different and innovative training methods. The US Army is meeting this challenge through increased use of simulators and simulations.

However, regardless of how inexpensive a simulator or simulation may be, it must ultimately enhance performance on the actual equipment and task being simulated. Kraemer & Wong (1992) showed that performance in a tank platoon gunnery simulator improved over exercises run. Demonstrating improved performance on the training device is important. However, for tank and BFV platoon gunnery simulators, the critical research question concerns whether degree of use and/or proficiency on platoon gunnery simulators relates to live fire platoon gunnery performance.

Platoon Gunnery Trainers (PGTs) for US Army tank and BFV platoons consist of four linked individual crew trainers called Unit Conduct of Fire Trainers or U-COFTs. These U-COFTs are virtual simulations of the commander and gunner crew positions. An instructor operator (I/O) plays the role of driver (and loader, for the tank), following the commander's instructions.

Platoons train using established exercises, such as hasty defense or offense. Platoons move over a pre-determined route (although they can vary speed) and engage targets that appear in a pattern standardized for each separate exercise. In addition to training gunnery, the simulator also trains command and control and fire distribution.

Platoon leaders must give movement and fire commands. The simulators can also train fire discipline and section gunnery. For instance, platoons must have some method of determining

which crews engage which targets to avoid either duplicating engagements or failing to engage certain targets.

Method

This research is correlational versus experimental. Our guidance was that any data collection had to be "transparent" to the units; that is, totally unobtrusive. Therefore, we could not randomly assign different platoons to different conditions, but could merely record the PGT performance of the platoons.

Sample

We collected PGT and live fire data on 35 M1A1 tank and 36 BFV platoons from US Army Europe (USAREUR).

Measures

We constructed two measures of PGT performance; total exercises run and total exercises passed. A platoon passed an exercise if it achieved a score of 70 or higher. The score roughly reflected the percentage of total targets killed. We constructed this information directly from electronic records made by the simulator. For tank platoons, we used the total exercises performed between gunnery rotations (platoons fire gunnery semi-annually). For Bradley platoons we used only the exercises performed within the same quarter as the current gunnery. However, both types of platoons completed most PGT training within a few weeks prior to gunnery.

We used two types of platoon gunnery or Table XII (TXII) measures; targets killed/targets presented and targets killed/targets represented. Tank TXII (TTXII) used what the Army calls a depleting scenario for main gun targets. For example, on the first band one might present 14 targets. If the platoon hit 8 targets on the first band, one would present only the 6 remaining targets on the subsequent, closer band. If the platoon killed the remaining 6 on the second band, its score in terms of targets presented would be 14 targets killed/20 targets presented (14 on the first band and 6 on the second), or 70 percent. However its score in terms of targets represented would be 14/14 or 100 percent, since the targets represented 14 enemy vehicles advancing toward them.

Bradley TXII (BTXII) did not use a depleting scenario. For example if the platoon hit 8 of the 10 targets on the first band, one would present all 10 on the second band anyway. If the platoon hit 2 of these targets the platoon's score in terms of targets presented would be 10 targets hit over 20 presented or 50 percent. The score in terms of targets represented would 10 out of 10 or 100 percent. In rare instances where a platoon hit more targets than targets represented, we recorded the score as 100 percent.

We constructed tank platoon performance for total targets, main gun targets, and troop (machine gun) targets. BFV platoons could engage troop targets with dismounted troops as well

as the BFV. Since the Bradley PGT did not train dismounted troops, we limited Bradley data collection to main gun targets.

Personnel belonging to USAREUR, not the units being trained, collected all TXII data used in these analyses.

Based on subject matter expert (SME) recommendations and prior research, we computed TTXII performance using current gunnery performance - prior gunnery performance. We defined BTXII performance as current gunnery performance only. SME rationale for the tank performance measure was that the PGT was particularly effective in sustaining tank platoon performance.

Results

Tank Platoons

We found that PGT use and proficiency related to live fire performance. Correlations displayed in Table 1 show that number of PGT exercises run and passed between gunnery rotations related to changes in gunnery performance between gunnery rotations. The more exercises run and passed, the more positive the change in percentage of total and main gun targets killed. Figure 1 shows the relationship between PGT exercises passed and change in percentage of total targets presented that were killed.

BFV Platoons

PGT use related to live fire performance. Correlations show that number of PGT exercises run prior to gunnery related to both measures of TXII performance. The more PGT exercises run prior to gunnery, the more main gun targets killed. Figure 2 shows the relationship between PGT exercises run and percentage of main gun targets represented that were killed.

Table 1

PGT Use & Proficiency and TXII Performance

	TTXII Targets (n = 35)					BTXII Targets (n = 36)	
	Total		Main Gun		Troop	Main Gun	
PGT Exercises	Rep	Pre	Rep	Pre	Pre	Rep	Pre
Run	.41*	.21	.51*	.30	-.16	.39*	.36*
Passed	.63*	.34*	.65*	.35*	.11	-.08	-.01

$p < .05$, two-tailed test

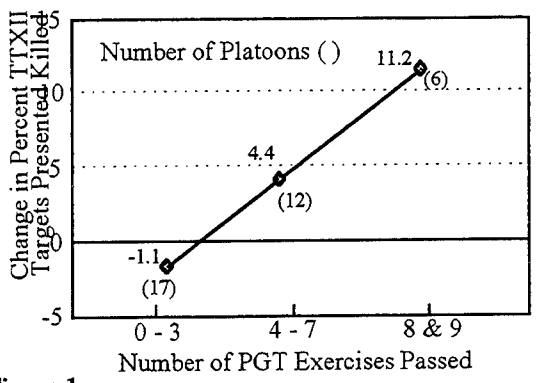


Figure 1.

PGT EXERCISES PASSED BETWEEN GUNNERIES AND CHANGE IN TTXII PERFORMANCE

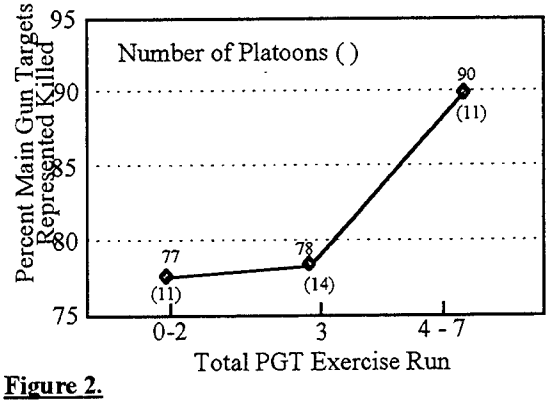


Figure 2.

PGT EXERCISES RUN AND BTXII MAIN GUN PERFORMANCE

Discussion

These results replicate previous findings for both tank (Sterling, 1993a) and BFV (Sterling, 1993b) PGT training. Although the data are correlational in nature, their replication suggests that PGT provides appropriate training for live fire gunnery. These results also demonstrate the utility of maintaining a training data base with both simulator data and data reflecting performance on the actual weapon system. Researchers can use the data to explore relationships between simulator use/proficiency and performance on the actual system, and how those relationships may change over time. Also, the data base may help to provide information to decision makers concerning the optimum amount of simulator training for a given level of performance on the actual system.

References

Kraemer, R.E., & Wong, D.T. (1992). Evaluation of a prototype Platoon Gunnery Trainer for Armor Officer Basic Course training. (ARI Research Report 1620). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A254 289)

Sterling, B.S. (1993a). Impact of tank PGT training on successive Tank Table XII scores (Training Note # 7). Grafenwoehr, Germany: Seventh Army Training Command.

Sterling, B.S. (1993b). Relationship between M2/3 Platoon Gunnery Trainer (PGT) training and platoon gunnery performance (Training Note # 12). Grafenwoehr, Germany: Seventh Army Training Command.

Psychomotor Abilities

Panel

Patrick C. Kyllonen
Scott R. Chaiken
Joshua B. Hurwitz
Lee Gugerty
Air Force Armstrong Laboratory

Abstract

The purpose of the panel is to discuss ideas and present data concerning the nature, organization, and measurement of human perceptual and motor abilities. A long-term goal of the research is to develop tests of human perceptual and motor abilities that might some day be used in a personnel selection and classification system to select pilots, air-traffic controllers, and other "real-time" human operators. The panel will discuss research concerning what psychomotor abilities are, how we best can measure them, how they are related to one another and to cognitive abilities, how they are affected by stressful conditions, and how they determine performance on complex tasks such as piloting an aircraft. (The panel will present several computer demonstrations of various tasks developed.) The key feature of the work is that it represents a synthesis of traditional psychometric approaches to researching human psychomotor abilities (e.g., Fleishman) with a cognitive, information-processing approach based on the Air Force's Learning Abilities Measurement Program's (LAMP) Cognitive Abilities Measurement (CAM) framework and family of associated cognitive-psychometric models. The panelists propose that a synthesis of these two approaches, the cognitive and the psychometric, will result in a new, more detailed specification of psychomotor abilities, and that this new specification will enable a more efficient system for measuring such abilities in operators and trainees.

Psychomotor Abilities

Panel

Patrick C. Kyllonen
Scott R. Chaiken
Joshua B. Hurwitz
Leo Gugerty
Air Force Armstrong Laboratory

Panel members have taken several different approaches to examining psychomotor abilities. These will be discussed in the form of different activities.

Activity I. We have conducted a series of exploratory factor-analytic studies that yield dimensions underlying performance on psychomotor tasks. Such studies have been done before (Fleishman, 1964) and have resulted in the suggestion that there may be eleven or so psychomotor factors (e.g., reaction time, multi-limb coordination, response orientation). However, what has not yet been attempted, is a study in which a wide range of both cognitive and psychomotor tests has been administered together. We will review the primary outcome of these exploratory studies--a specification of a set of psychomotor factors and their interrelationships to each other and to cognitive factors.

Activity II. We have conducted a series of confirmatory tests of models derived from the Psychomotor-Cognitive Abilities Measurement (PCAM) taxonomy. The PCAM taxonomy specifies process rows (e.g., working memory, declarative learning) and domain columns (e.g., verbal, spatial) as dimensions underlying test performance. To accommodate psychomotor tasks, the taxonomy posits temporal and motor aspects of tasks as constituting additional task domains (i.e., additional columns).

Activity III. We have begun exploring the possibility that stressors, such as time pressure, or decision risk, add a new dimension to the PCAM taxonomy. The necessity of a new dimension is indicated in two ways. First, for every PCAM task, it should be possible to create a counterpart through the additional imposition of a stressor. That is, the stressor attribute is simply an additive feature of the task's definition. For example, a temporal working memory task could be administered both with and without a stressor. Second, stressor versions of tasks should show empirical independence from non-stressor versions.

Activity IV. We have begun evaluating the sufficiency of models derived from the PCAM taxonomy by creating a PCAM test battery and using that battery to predict performance in a variety of complex, real-time, multi-tasking situations. These include (a) a driving simulator that emphasizes situational awareness, (b) a basic flight instruction tutor and simulator that instructs students on how to fly a Cessna 152, (c) an accompanying system that measures performance on a real aircraft, and (d) a simulated air combat "situational awareness" measurement system.

Behavioral Indicators of Effective Performance and Leadership as Identified Through a Policy-Capturing Method

Captain Linda S. Hurry
Headquarters, USAF Air Combat Command
Guy S. Shane, Ph.D
Lieutenant Colonel James R. Van Scotter, Ph.D
USAF Institute of Technology

Abstract

The study used policy capturing (Hobson & Gibson, 1983) to test the contribution of four types of work behavior to the effectiveness of junior Air Force officers as judged by Air Force supervisors. Results showed that all four types of behavior (leadership, task performance, interpersonal facilitation and job dedication) each contributed significantly and independently to overall performance. This finding held across job specialty, and was not affected by supervisors' demographic characteristics except for rank. Field grade officers and company grade officers gave different weights to task performance and interpersonal facilitation.

At least since Sun Tzu wrote of the qualities required of the commander (c.400 BCE/1963), military leaders have been investigating the characteristics that are associated with officer effectiveness. The demise of the Cold War and other environmental changes have caused the services to readdress the issue of officer effectiveness. The US Air Force has particularly been interested in adapting educational and training programs for new officers to meet changing performance requirements.

This study tests a model of officer effectiveness suggested by recent research. Our objective is guide Air Force efforts to improve the relevance of commissioning training programs by defining junior officer performance requirements in terms of the specific behaviors, skills, and characteristics that separate highly effective officers from less effective officers. This behavioral focus is consistent with current theories of learning (Kirkpatrick & Locke, 1991, p. 501.) and with the need to provide detailed and practical guidance to the USAF commissioning sources (the Air Force Academy, Officer Training School, and the Reserve Officer Training School).

Leadership is a requirement for military success. Recent research (ACSC, 1988; Bausum, 1986; Borman & Motowidlo, 1993; Borman & Brush, 1993; and Van Scotter & Shane, 1995) has described leadership in behavioral terms. Similarly, research to define performance has produced evidence that supervisors expect subordinates to perform in a variety of ways that go beyond narrow job descriptions (Van Scotter, 1994, p.88). For example, behaviors such as working hard, persisting, taking initiative, and paying attention to details differ from job-specific task performance, but are clearly important in most jobs. Borman and Motowidlo (1993) argued that in performing contextual behaviors like these employees contribute to the effectiveness of the

organization supporting an interpersonal and social climate that supports mission accomplishment. Through multiple regression analysis, Motowidlo and Van Scotter showed that both task performance and contextual performance were "uniquely and significantly associated with overall performance" (1994, p. 479) as measured by supervisory performance ratings. Factor analyses by Smith, Organ and Near (1983), and by Motowidlo, Packard and Manning (1986) each indicated that there were two distinct factors within the contextual domain. Van Scotter (1994) divided the contextual domain into interpersonal facilitation and job dedication and found that each category contributed uniquely to the effectiveness of Air Force maintenance technicians. We therefore included two contextual variables along with the task dimension in our model. Thus, a model of effective junior officer performance should include task behaviors, contextual performance behaviors, and leadership behaviors.

Policy capturing determines the importance of various decision variables to the choices made by participants familiar with the situation. Respondents are presented with multiple scenarios that vary the factors of interest to the researcher. At the end of each scenario, respondents make choices (Webster & Trevino, 1995). Policy capturing thus infers the importance of factors from individuals' actual choices rather than from their reports of how they make decisions. The method eliminates one source of error from subjective ratings of their decision-making priorities, and results in a construct-valid representation of "true" rating policies (Hobson & Gibson, 1983, p. 641). Thus, policy capturing can be an accurate simulation of the judge's decisions (Hobson & Gibson, 1983).

Method

Our research showed that overall performance ratings of officers should be explained by behaviors from four behavior domains or factors. These are: leadership, task performance, interpersonal facilitation, and job dedication. Each of these constructs has been shown to account for significant and unique variance in overall performance in related studies (e.g., Borman & Motowidlo, 1993; Van Scotter, 1994). We hypothesized that they would apply equally well to junior Air Force officer performance.

Instrument Development. We developed the rating instrument by creating a set of behavioral descriptions each designed to represent one of the four behavior domains. Seventy-nine Air Force officers in the ranks of Captain through Colonel and five Lieutenants with an average of 6.5 years' experience in supervising junior officers rated the importance of 102 candidate behavioral descriptions. These were subjected to reliability analysis using Cronbach's alpha (see Table 1) and principal components analysis to refine the fit to the dimensional categories. These were further refined using an item intercorrelation analysis suggested by W. H. Hendrix (personal communication, 10 April, 1995). This resulted in a final set of items in each behavioral dimension as shown in Table 1.

Table 1
Behavioral Dimensions Resulting from Principal Components Analysis

Dimension	Reliability	Number of Items
Leadership	.91	15
Task Performance	.89	13
Interpersonal Facilitation	.93	18
Job Dedication	.92	18

Notes: Total N of items = 64

Reliabilities computed using Cronbach's alpha Respondents. Respondents were 210 Air Force officers stationed at Wright-Patterson Air Force Base, chosen based on being currently assigned as a supervisor of junior officers. Of these; 26 were rated, 130 support, 45 analysts/engineers; additionally, there were 171 males and 30 females and 138 company grade and 63 in field grades. Nine questionnaires were discarded because of unusable data, leaving 201 analyzable sets of responses.

Procedure. The rating instrument was computerized by one of the authors (Van Scotter) for ease of administration and reliability as well as initial data compilation and analysis. The questionnaire asked for demographic information, explained the performance categories, established a scenario for the respondents, and explained the information in each profile as well as one practice profile. They then responded to profiles of 50 hypothetical junior officers each of which consisted of combinations of four rated behaviors from each of the performance dimensions. Raters viewed the hypothetical "typical job performance" in each profile and then provided a rating of overall performance on a scale of 1-5.

Analysis. Multiple regression was used to test the hypothesis that each behavior dimension, job dedication, interpersonal facilitation and leadership, explained a unique portion of junior officer performance, the dependent variable, operationalized as the overall performance score. Regressions were computed for each respondent (the rating policies) which were then combined to derive the policy for the entire sample. Mean beta weights estimate the contribution of each behavioral dimension to overall performance.

Analysis of variance (ANOVA) was accomplished including Tukey's procedure for multiple comparisons using a .05 level of significance. We included tests of differential effects by job type of the respondent and by demographic category.

Results

Intercorrelations of the behaviors that comprised the 50 profiles are shown in Table 2. They provide evidence that the dimensions are relatively independent. Cronbach's alphas indicate responses to the leadership, task performance, interpersonal facilitation, and job dedication items were consistent.

Table 2
Intercorrelations among the performance dimensions

Dimension	1	2	3	4
1. Leadership	(.91)	---	---	---
2. Task Performance	.04	(.89)	---	---
3. Interpersonal Facilitation	-.14	.03	(.93)	---
4. Job Dedication	-.04	.05	-.13	(.92)

Notes: N = 50 profiles. 201 Supervisors. $p < .05$ for $r > .037$.
Alpha reliabilities are shown on the diagonal.

The relationships between leadership, task performance, interpersonal facilitation, job dedication as independent variables and supervisors' overall performance ratings as the dependent variable were tested via multiple linear regression. A total of 201 sets of overall ratings based on the 50 hypothetical profiles were regressed on item mean scores (developed in the preliminary research) for each of the performance categories. The mean beta weight for leadership was β = .46 (T = 19.19), task performance was β= .34 (T = 19.96), interpersonal facilitation was β = .27 (T = 15.07), and job dedication was β = .17 (T = 25.44). All significantly different from zero ($p <$ .05), indicating that each performance dimension contributed significantly to the supervisors' overall judgments.

Performance dimension by occupation group (4X3) ANOVA analyses were conducted to test the possibility that occupational groups viewed performance differently. Results showed that performance category had a significant influence on the ratings ($F=100.33$, $df=2$, $p<.01$), but occupation, and an occupation by performance dimension interaction term did not. Those results coupled with the standardized regression weights shown in Table 3 indicate a remarkable degree of consistency in the supervisors' judgments across occupational areas.

Table 3
Standardized regression weights for supervisors from three occupational groups

	Officer Job Type			
Behavior Category	Support	Rated	Anal/Eng	Total
Leadership	0.47	0.41	0.45	0.46
Task Performance	0.34	0.39	0.31	0.34
Interpersonal Facilitation	0.26	0.24	0.31	0.27
Job Dedication	0.16	0.17	0.19	0.17

Notes: Total N=201. N=130 support officers; N=26 rated officers; and N=45 analysts/engineers.

All standardized beta weights significant ($p<.01$).

Additional ANOVA analyses tested the effects of grade, race, sex, and the supervisor's commissioning source on his or her rating policy. Of these factors, only the raters grade significantly influenced the rating policy. The interaction between grade and performance category was significant ($F_{137,62} = 3.783$, $p < .05$). Follow-up analyses showed that company grade and field grade officers differed in the weights they gave to task performance ($F_{137,62} = 3.62$, $p < .05$) and interpersonal facilitation ($F_{137,62} = 6.84$, $p < .05$), with company grade officers rating interpersonal facilitation ($Q = 2.902$, $p < .05$) as more important, and field grade officers rating task performance ($Q = 2.902$, $p < .05$) as more important.

Discussion

This research supported a four-factor model of junior officer effectiveness. The data indicate that junior officer performance involves a mixture of behaviors from at least four areas -- leadership, task performance, interpersonal facilitation, and job dedication. The data showed that the importance of the four categories varied little across officer groups formed by occupation, gender, race, grade, or commissioning source. The results have several important implications. The findings concerning occupational differences indicates that the commissioning sources do not have to create separate curricula for different types of jobs. It is important to note that the leadership category is the most important contributor to overall performance regardless of officer job category. This suggests the commissioning sources should continue placing emphasis on leadership and also consider increasing the amount of instruction on leadership behaviors useful in real management situations.

Finding that rating policies varied little across demographic groups suggests supervisors use similar criteria regardless of their personal background. The commissioning source, race, and sex of the rater did not affect the importance of the performance categories in the overall performance evaluation. The grade of the rater, however, played an important role in determining the impact of task performance and interpersonal facilitation. Field grade officers placed greater importance on task performance than did the company grade officers; whereas the company grade officer rated interpersonal skills as more important for effective performance than the field grade officers. With longitudinal data, it might be possible to determine whether this is because officers change their views on performance as they achieve higher ranks, or perhaps officers who emphasize task performance are more likely to achieve higher ranks. Since my data were cross-sectional, the effects of other, possibly unmeasured, variables can not be ruled out. Further research investigating differences with the way field grade and company grade officers view performance may lead to improvements in training that might shorten the learning curve.

References

Air Command & Staff College (ACSC) Students (1988, September). Guidelines for command. AU-2. Maxwell AFB AL: Air University Press.

Bausum, H. S. (1986). The John Biggs Cincinnati lectures in military leadership and command 1986. Lexington, VA: VMI Foundation, Inc.

Borman, W. C., & Brush, D. H. (1993). More progress toward a taxonomy of managerial performance requirements. Human Performance, 6, 1-21.

Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. W. Schmitt & W. C. Borman (Eds.), Personnel selection in organizations. New York: Jossey-Bass.

Hobson, C. J., & Gibson, F. W. (1983). Policy capturing as an approach to understanding and improving performance appraisal: A review of the literature. Academy of Management Review, 8, 640-649.

Kirkpatrick, S. A., & Locke, E. A. (1991) Leadership: Do traits matter? The Academy of Management Executive, 5, 48-60.

Motowidlo, S. J., Packard, J. S., & Manning, M. R. (1986). Occupational stress: Its Causes and consequences for job performance. Journal of Applied Psychology, 71, 618-629.

Smith, N. W., Organ, D. W., & Near, J. P. (1983). Organizational citizenship behavior: Its nature and Antecedents. Journal of Applied Psychology, 68, 453-463.

Sun Tzu. (1963) The Art of War (S. B. Griffith, Trans.). New York: Oxford University Press. (Original work published c. 400 BCE)

Van Scotter, J. R. (1994). Evidence for the usefulness of task performance, job dedication, and interpersonal facilitation as components of performance. Unpublished doctoral dissertation, University of Florida, Gainesville, FL.

Van Scotter, J. R. & Shane, G. S., (1995). Interrelationships among properties of performance rating criteria and their relationships with supervisor experience and training. In S. B. Rasch, (Ed.), Proceedings of the Human Resource Management Group, Association of Management: 13th Annual International Conference, 13. (pp 112-127), Vancouver, BC, Canada: Maxmilian Press.

Webster, J., & Trevino, L.K. (1995). Rational and social theories as complementary explanations of communication media choices: Two policy-capturing studies. Academy of Management Journal, 38, 6, 1544-1572.

Cognitive Therapies for Intelligent Organizations

John R. Landry, Ph.D.
Metropolitan State College of Denver
CMS Dept., School of Business
Denver CO 80217-3362
landryj@mscd.edu

Abstract

Organizations are increasingly viewed as products of their members' cognitions that create organizational memories and routines through enactment and learning. One relatively unexplored aspect of this view is the creation of organizational psychoses. This paper adapts a model of cognitive therapy to organizations, illustrates the model's applicability by citing contemporary problems of some well-known businesses, and suggests how organizations create conditions that may generate and sustain these patterns of behavior.

The malady of commercial crises is not, in essence, a matter of the purse but of the mind.
John Stuart Mill, 1867

Among the metaphors used to describe an organization, the organization as a mind is an intriguing approach, particularly when organizations are dominated by knowledge workers. Phenomena such as organizational memory and organizational learning are widely discussed and debated. However as organizations are less often thought of as machine bureaucracies and more frequently as cognitive structures, it is reasonable to assume that organizations may exhibit symptoms of psychosis (Morgan, 1986). While the approaches of Kets De Vries and Miller (1982), Schaef and Fasse (1988), and Feinberg and Tarrant (1995) emphasize individuals and their mostly unconscious processes in creating distressing organizational behaviors, this paper's approach begins with an observed behavior and then traces it to inappropriate thoughts. The purpose of this paper is to suggest how the organizational "mind" responds to disturbances.

A Cognitive Model

The work of Beck and Ellis has become closely associated with cognitive therapies to aid individuals (Santrock, 1988). Among their major contributions is the identification of a small set of responses to disturbances that are found across the population. Their therapies are based upon the notion of identifying the disturbing responses and actively refuting the implied consequences of the evoked responses. Providing managers with a similar framework may mitigate the need for external parties (e.g., consultants) to intervene in organizational processes.

Freeman and DeWolf's (1992) book, The 10 Dumbest Mistakes Smart People Make and How to Avoid Them, identifies inappropriate individual thoughts. Their framework is adopted for this paper and extended to organizations through analogy. The basis for using an analogy can be

found in Argyris and Schön's (1978:12) observation that organizations are a "cognitive artifact." Since the cognitive constructionist view centers on members' cognitions, it is not unreasonable to assume that members' cognitions may also lead to organizational psychoses. Morgan (1986:202) asserts that "organizations and their members can become enmeshed in cognitive traps. False assumptions, taken-for-granted beliefs, unquestioned operating rules, and numerous premises and practices can combine to create self-contained views of the world that provide both a resource for and a constraint on organized action." Freeman and DeWolf (1992:xxii-xxiii) express it this way: "The situations may vary, but the point is the same: Different thoughts produce different emotions."

It is the thinking individuals with their emotions, actions, and thoughts that constitute the organization. Freeman and DeWolf (1992) list ten thinking patterns that lead to dysfunctional behaviors. Although each pattern can be described in a sentence or two, Freeman and DeWolf use short labels that are memorable. It is their sometimes whimsical labeling that makes their scheme appealing for translation into the organizational realm since it provides managers a way to quickly communicate with each other. Table 1. briefly introduces the cognition labels and provides a description.

Pattern	Description
"Chicken Little Syndrome"	Inferring catastrophic conclusions possibly resulting in paralysis
"Mind Reading"	Believing that you know what others think and they should know what you think
"Personalizing"	Taking responsibility for others' circumstance or a natural phenomenon
"Believing Your Press Agent"	Generalizing successes in one area to other areas without the same effort
"Inventing and Believing Critics"	Unquestioning acceptance of others' criticism or finding criticism from others
"Perfectionism"	Demanding perfection in all areas or setting impossibly high standards
"Comparisonitis"	Focusing on comparisons yielding negative evaluations or accepting negative comparisons by others
"What If" Thinking"	Worrying about non-existent or improbable events or worrying excessively about real threats
"The Imperative 'Should'"	Intensely focusing on a mandate that may not be refused or past refusals for imperatives
"Yes-Butism"	Providing negatives that outweigh positives or concocts reasons to dismiss an obvious negative

Table 1. Descriptions of Ten Potentially Dysfunctional Thinking Patterns of Individuals Adapted from Freeman and DeWolf (1992:12-13)

One caveat to the application of this framework warrants attention before getting into the details. The framework proposes a set of responses that are, for the most part, completely

appropriate but carried to an extreme they become dysfunctional. It is the exaggeration of the response that becomes a concern rather than the response itself.

Mistakes Organizations Make

Since this paper focuses on the organization rather than the individual, group behaviors are more relevant than the actions of a single individual. These behaviors, and their negative consequences, are illustrated by using recently published articles that describe well-known organizations who are encountering difficulties -- Rubbermaid, TRW and Apple Computer.

For a decade Rubbermaid has been in the top ten of Fortune magazine's most admired companies. With this outstanding reputation, Rubbermaid has faltered recently. "But in the 1990s, Rubbermaid has had to struggle a lot harder for growth. In fact, it is falling short. Even though the target remains 15%, sales grew only 8% in 1993 and 10% in 1994. Earnings did grow 15% in 1993, but only 8% in 1994. Inflation, which once boosted results 5% or so annually, can now be relied on for no more than 2% or 3%. (Smith, 1995: 92)" Their comparison uses a tough standard—their own past performance— a self-induced Comparisonitis.

While Rubbermaid's goals are arguably too high, they have nevertheless maintained those goals and angered customers in the process. The biggest customer was Wal-Mart. "To meet sales quotas, lots of product is dumped at very deep discounts at the end of the quarter. Rubbermaid's monitoring system, designed to keep track of the complex tangle of purchase orders the company negotiates with thousands of customers, apparently failed to detect that someone was getting a phenomenal deal—and that Wal-Mart wasn't. (Smith, 1995:100)" To the sales reps, the sky was falling at the end of the quarter and The Chicken Little Syndrome kicks the panic button that has far-reaching consequences.

Another well-known organization, TRW, manufactures, among other things, automotive air-bags. The main manufacturing site has been plagued with mishaps and was recently closed by fire officials for several days after numerous responses to alarms. "When a worker waited nearly an hour for an ambulance after accidentally backing a forklift off a 5-ft. platform, some employees say they were worried managers were reluctant to call for help out of fear of attracting more media attention. (Schiller and Schine, 1995:63)" "What If' Thinking lead to potentially serious consequences for the worker and appropriate action was blocked by focusing on the worst possible public interpretation of the event.

In addition to the forced plant closing, the public's attention was focused on TRW by a $1.7 million dollar fine from a legal action related to the death of a worker and the injury of another. Another fine of $89,000 was assessed by the state's occupational health office for ineffective fire-prevention and protection programs. Both of the actions were associated with company procedures. TRW may be responding by Personalizing every criticism leveled at them since they have had a string of unfortunate events. John Janitz, executive vice-president of the TRW group than oversees the plant, and apparently with the support of others "argues that the intense scrutiny is unfair, and he questions whether TRW is being measured by the same yardstick as rivals. Company supporters wonder if fires and explosions at other manufacturers' more remote plants are simply going unnoticed. (Schiller and Schine, 1995:63)"

The recent fortunes of Apple Computer have been less than bright, partially because they ignored an eminent threat -- Microsoft's Windows. "But when the first commercially successful version of the program appeared in 1990, Apple's initial reaction was dismissive. The day Windows 3.0 was launched, Apple's executive staff—including Sculley—gathered for a demo. 'They were mocking it,' said a former Apple manager who was there. 'They said it was awkward, clumsy, a piece of junk. They were laughing. It was complete arrogance.' (Rebello, Barrows and Sager, 1996:40). Although Apple was highly rated at the time, their actions were consistent with Believing Your Press Agent and ignoring latent threats.

But that was not Apple's only belief that had deleterious effects. "Insiders say that Ian Diery, now president of AST Research, Inc., prepared a 1995 forecast of about 15% growth—the same as analysts were predicting for the industry. But CFO Graziano argued that was not aggressive enough if Apple still hoped to expand market share. The combative Diery insisted on a less ambitious agenda, according to insiders, saying the company hadn't hit its plan in years, and employees need to 'feel like they could win.' ... Diery won. (Rebello et al., 1996:40)" Apple may have become their own worst enemy by Inventing and Believing Critics.

Apple's Perfectionism was deeply embedded in their culture. "There has always been a flip side to the Apple ethos, though. 'The culture has incredibly powerful elements—Jobs's perfectionism, for one,' says a former Apple executive. ... Inevitably, that lead to clashes between the 'creators,' such as Jobs and his Mac mates, and the experienced managers hired to run marketing and finance. (Rebello et al., 1996:39)" When John Scully took over as CEO he signaled his support for valuing perfectionism by "lionizing the technical 'wizards (as they described themselves on business cards), ... (Rebello et al., 1996:39)"

As the fortunes of the company changed they received an Imperative "Should" to gain market share (Rebello et al., 1996:41) "..., he [Michael Spindler, Apple's CEO] ordered an all-out bid for market share." The organization's response was to engage in Mind Reading: "Apple marketing execs had misread consumers: Apple had too many low-end models and too few of the powerhouse that buyers were snapping up. When the [Christmas] wrapping paper settled, Apple was left with $80 million worth of inventory write-offs, while IBM, Compaq Computer, and HP had cleaned up. (Rebello et al., 1996:41)"

Finally, Apple has been wrestling for years with a decision to license their operating system. In mid-1994, they adopted a plan to allow licensing (Rebello et al., 1996). After this apparent agonizing decision process yielded an outcome many industry observers think appropriate, Apple's own executives could not agree on who qualified for the plan. Their reasoning was a concern for market erosion although it may be an implicit perception that they will suffer a loss of control. What appeared to be a victory was diminished by the Yes-Butism that put forth reasons why it would not, could not, or should not work — defeat snatched from the jaws of victory!

Discussion

While these examples cannot offer substantive support for this paper's thesis, they are illustrative of organizational behaviors that may be driven from faulty premises. Organizations can become sensitive to these inappropriate behaviors and begin processes to discover and refute

the underlying assumptions. Participants also showed some awareness of their angst, emotional distress or physical danger. These very personal signals should not be ignored and they should provoke further investigation. Also, the behaviors are not inappropriate until, when taken to an extreme, get in the way of effective organizational coping. Additionally, each organization appeared to have more than one inappropriate behavior that in combination exacerbated the situation. Moreover, it is not unreasonable to posit that some normal organization processes may lead to ineffective coping and these activities deserve special attention (Table 2.).

Pattern	Organizational Activities
Chicken Little Syndrome	Worst-case scenarios, disaster drills and task forces
Mind Reading	Competitor analysis, market research and forecasting
Personalizing	Customer surveys, focus groups and consumer ratings
Believing Your Press Agent	Merger-acquisition, new offerings and mission statements
Inventing and Believing Critics	Audits, investment reports and credit ratings.
Perfectionism	Just-in-time, TQM and zero-defect programs
Comparisonitis	Benchmarking, process control, variance reports and goals
"What If" Thinking	Scenario building, risk assessments, and simulations
The Imperative "Should"	Policies, procedures, and "routine" meetings and reports
Yes-Butism	Devil's advocacy, status meetings and project justifications

Table 2. Speculations about activities that may bolster inappropriate organizational responses

Although Table 2. is speculative and certainly not comprehensive, organizations need to pay heed to the processes that they create by attending to the assumptions, belief, and ideals enacted.

References

Argyris, C. & Schön, D. (1978). Organizational learning: A theory of action perspective, Addison Wesley, Reading, MA.

Feinberg, M. J. & Tarrant, J. J. (1995). Why smart people do dumb things, Fireside, New York, NY.

Freeman, A. & DeWolf, R. (1992). The 10 dumbest mistakes smart people make and how to avoid them, Harper Collins, New York, NY.

Kets De Vries, M. F. R. & Miller, D. (1984). "Neurotic style and organizational pathology," Strategic management journal, Vol. 5, pp. 35-55.

Morgan, G. (1986). Images of organization, Sage, Beverly Hills, CA.

Rebello, K., Burrows, P. & Sager, I. (1996). "The fall of an American icon: Apple Computer, once the hip flagbearer of high tech, is in sad decline. There are lessons aplenty," Business Week, Feb. 5th., 34-42.

Santrock, J. A. (1988). Psychology: The science of mind and behavior, 2nd. Ed., Wm. C. Brown, Dubuque, IA.

Schaef, A. W. & Fassel, D. (1988). The addictive organization, Harper & Row, New York, NY.

Schiller, Z. & Schine, E. (1995). "An explosive mix at TRW: Air-bag-plant mishaps—and bad press—haunt the company," Business Week, Dec. 18th., 62-63.

Smith, L. (1995). "Rubbermaid goes thump," Fortune, Oct. 2nd, 90-104.

Reengineering the Human Interface with Space: a Team Approach to Process Improvement

Lt Col Frank McIntire
HQ Air Force Space Command
Mr Chip Houlihan
KPMG Peat Marwick LLP

Abstract

Lightning-fast results with no barriers standing in the way describe Air Force Space Command's newest organizational improvement strategy known as Business Process Reengineering. In one short week, the team-based initiative allowed process owners to reduce Peacekeeper missile stage three processing cycle time by 78 percent, reduce distance traveled by 71 percent, and reduce processing man-hours by 72 percent. This high-energy event involved 95 percent action and 5 percent debate. On-the-spot changes were made to inefficient procedures that have frustrated and confounded willing workers for ten years or more. By assembling the right team and by using the right model, the command cut through the red tape like a laser beam.

Air Force Space Command has experienced tremendous success in its Quality Air Force endeavor. It is important to note that Space Command's initiatives are consistent with other Federal and Department of Defense programs designed to create a government that works better and costs less. Some of these include:

- Chief Financial Officer's Act
- Government Performance and Results Act
- Weapon Systems Cost Reduction
- Integrated Weapon System Management
- Acquisition Reform

Major initiatives include the use of Quality Councils, QAF Training, Improvement Teams, Organizational Planning, and Unit Self Assessment. Within the realm of Improvement Teams, Space Command in the past has relied on ad hoc tiger teams, natural working groups with a wide array of structures, and the process action team that utilizes the seven-step Continuous Improvement Process. Through its relationship with Air Force Materiel Command (AFMC), Space Command was introduced to an excellent example of another approach to team-based process improvement: a total redesign of the process from the ground up. The Space Systems Support Group, an AFMC unit assigned to Peterson Air Force Base discovered that its Emergency Depot Level Maintenance Process was hopelessly broken and needed much more than improvement; it required a radical redesign. They applied what has come to be known as Business Process Reengineering to that process. It was this same Business Process Reengineering that was adopted by Space Command to revamp the process for performing heavy maintenance on the Peacekeeper missile. Through Business Process Reengineering, the command was successful in creating a completely new process for missile stage processing that saves a

tremendous amount of time, a modest amount of money, and eliminates irritating bottlenecks and red tape which have been in place since the missile was fielded.

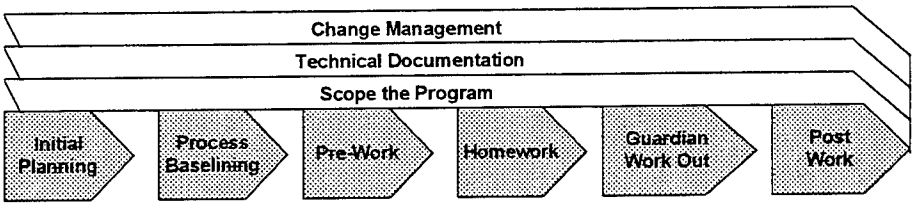
This new level of innovation and team-based improvement called Business Process Reengineering is now part of Space Command's repertoire for team-based process improvement. The methodology has been given the name "Guardian Workout". It offers a new level of sophistication for Improvement Teams, supports the overall goal of reducing the cost of operating the government, and is the subject of this paper.

Method

Heavy maintenance on the Peacekeeper Missile is no understatement. The missile itself weighs-in at nearly 200,000 pounds, consisting of 4 booster stages and a warhead. The most striking difference between the Peacekeeper and the Minuteman from a maintenance perspective is that unlike the Minuteman which is handled like a single round of ammunition, the Peacekeeper is so massive that it must be handled stage by stage. It is shipped, received, processed, and placed into the silo where it is assembled one stage at a time. The reverse must occur when the missile is removed and returned to depot or shipped to Vandenberg for test-firing.

The processing of just one stage, in this case the third stage, has been known to take as long as three weeks. Rather than merely look for opportunities to improve one or two steps in the process, it was decided that the entire stage three missile processing function would be completely reengineered. The business process reengineering components that would become characteristic of Guardian Workout are as follows: A Strategy for Change Management (Plan to Plan), a Requirement for Technical Documentation (Plan to Plan), Scope the Program (Plan to Plan), Initial Planning, Process Baselining, Pre-Work, Homework, Guardian Workout, and Post Work.

An illustration of the model employed for this initiative:



The framework for business process reengineering included a number of assumptions and limitations. Relative to the Guardian Workout itself: safety would not be compromised, regulatory guidelines would be challenged, senior leaders would promote a free flow of ideas, the team would have a minimum of four days to reengineer the process, and personnel cuts would not be considered. In determining cost savings: capital investment was found to be excess or fully depreciated, manpower costs were based on the standard burden wage schedule, and maintenance costs were computed using an annual historical average.

The concept of wartime readiness is abstract indeed. Yet the system which has been designed to produce readiness is evidenced by tangible elements: highly trained personnel, one very large missile, a facility designed for heavy maintenance, 40 wheel transporters, and detailed procedures.

Participants

The majority of missile stage processing activity at the wing or group level is performed by a team of 5 or 6 technicians. There is a diverse list of stakeholders, however and all were included in the reengineering effort. Representatives from Headquarters Air Force Space Command (HQ AFSPC) would ensure that the participants would have access to necessary resources and would eliminate unnecessary red tape. Over-the-shoulder advice was offered by Air Combat Command and Pratt & Whitney, both of whom have experience with large-scale improvement and a desire to learn more about the same. Process facilitators were provided for the reengineering teams. Host wing stakeholders included Safety, Quality Assurance, Logistics, and Quality Improvement. Representatives from Vandenberg AFB and Ogden Air Logistics Center also manage the Peacekeeper missile, and were very interested in this initiative. Finally the Technical writers from Ogden Air Logistics Center were invited to attend in order to make real-time changes to the Technical Orders should that become necessary.

Equipment

The equipment associated with Peacekeeper missile stage processing is typical of what one would expect at a maintenance depot. Peacekeeper missile stages are processed at an industrial area that is located on an expanse of real estate that is geographically displaced from the base proper. The majority of maintenance activity occurs at a centrally located missile stage processing facility (MSPF). The MSPF is served by a dedicated rail line which allows cargo to be rolled directly adjacent to the facility. A high-lift crane adjacent to the MSPF facilitates the loading and unloading of missile stages packaged for transport. At a separate location, missile stage transporting vehicles are parked in close proximity to a collection of missile storage containers of various shapes and sizes. The 40 wheel, articulated vehicles are specifically designed to carry the containerized missile stages. Also geographically displaced are three separate missile stage storage facilities each of which can accommodate a number of missile stages packed in containers. An elaborate system of roadways connect the storage locations with the MSPF and the high-lift crane.

Inside the MSPF are the typical collection of tools related to the business of heavy missile maintenance. Toolboxes and tool racks are loaded with a variety of hand and power tools. Although the high-lift crane is electrically powered, movement of the missile stages within the MSPF is facilitated by air-powered motors to reduce the danger of stray voltage causing damage or catastrophe. Large ring-shaped carriage adapters are attached to the missile stages and facilitate their movement and storage.

All activity surrounding missile stage processing is governed by technical orders. The activity within the MSPF is typical of large-scale aerospace vehicle testing and servicing. Missile

stages are positioned and repositioned within the MSPF in order to perform a variety of functional checks on electrical systems and components. Access panels are opened and secured in order to facilitate tests and operational checks. Voltage checks are performed to ensure electrical continuity, signal and processor checks are performed on hardware which supports propulsion and guidance, and battery performance and reliability checks are accomplished. Activity that occurs outside the MSPF takes place over a wide expanse of real estate and is largely related to the movement and repositioning of transport vehicles, missile stages, missile stage containers (empty or filled). Key locations outside the MSPF serve as pick-up, drop-off, and transfer points for transport vehicles moving containerized missile stages.

Process

The key to successful accomplishment of the reengineering initiative is the cooperation of process owners, operators, and policy makers to eliminate the non-value added activity, or eliminate some process steps entirely. The critical factors necessary to ensure success are as follows:

- Key staff elements must participate
- Process owners and practitioners must participate
- Access to all required cost data and processes
- Access to subject matter experts
- Access to all technical documentation
- Participation by all agencies that are involved in the development and control of missile stage processing
- Open dialogue and information exchange between headquarters and process owners
- An objective and unbiased collection and presentation of data
- Senior officer support (AFSPC/CC/CV, Wing/CC, Group/Squadron/CC

The reengineering initiative involves three major steps: week-long prework for the core team, one day of training for all participants, and a week-long effort to achieve target goals. During the week-long prework stage, the host unit identifies the scope of the reengineering effort, process name, process owners, suppliers, customers, and unit representatives. The core team identifies objectives, problem areas, performance measures, and regulatory barriers. The process that will be reengineered will be "mapped" to identify inputs, controls, outputs, and mechanisms. Historical data is reviewed, equipment and support requirements are established, and the steps required to accomplish the process are "walked out" and documented in terms of time and motion. A training area and daily debrief area (for upcoming reengineering effort) are designated, and sub-team leaders are selected. The core team departs the unit to plan the reengineering effort and allow the host unit to prepare for the reengineering effort.

The core team returns and joins all of the reengineering team members at the host unit for one day of training which will be followed immediately by the reengineering effort. The training consists of a rapid-fire diet of reengineering principles and methods to include: eliminating the 7 types of waste, fool-proofing, pull production, reducing cycle and lead times, just-in-time delivery, reduction of inventory and work-in-progress, reducing required floor space, visual

control, machinery simplification, multi-skilling and job enrichment, and effective follow-on activity once the process has been reengineered.

The reengineering participants considered the full scope of activity surrounding Peacekeeper missile stage three processing. Every process step was scrutinized relative to missile stage transportation, missile stage maintenance for launch facility, and missile stage maintenance for depot. Factors considered in each of the above mentioned categories include: set-up time, cycle time, man-hours, walking distance, and processing steps. During the week-long reengineering effort waste was systematically eliminated from the process as it was radically changed.

Results

Documented savings in time, distance, and process steps are outlined in the tables.

TARGET PROGRESS REPORT
STAGE III TRANSPORTATION

27 Nov - 1 Dec	Monday	Tuesday	Wednesday	Thursday	Friday	Target
Set-Up Time (Hrs)	2.1	1.0	1.0	.6	.6	1.0 hrs .5 hrs*
Cycle Time (Hrs) (includes set-up)	19.1	5.2	3.0	2.8	2.3	4.8 hrs 2.5 hrs*
Man-hours	54.6	14.4	10.4	9.7	7.0	15 hrs 10 hrs*
Walking Distance (miles)	7.8	2.0	2.0	2.0	2.0	5.0 mi 2.0 mi*
Processing Steps	68	23	22	16	14	34 20*

NOTE: *Stretch Goal Set

Air Force Space Command has positioned itself to take the lead in several Quality Air Force initiatives. Philosophy and "quality speak" are being replaced with methods for fostering organizational improvement and leadership development. The components of process management are now considered under the category of system management. The basic tool of process flowcharting is becoming a subcomponent of process mapping. The following goals were established with the expectation that there was tremendous waste in the system: reduce transporter process steps by 50%, reduce stage processing steps by 30%, reduce transporter cycle time by 75%, and reduce stage processing cycle time by 50%.

The majority of the goals that were established were achieved during the first three days of the reengineering effort. The most significant savings was in the cycle time reduction from three days to one. Additional benefits were a nominal savings of $1500 per unit and a significant reduction in irritating red-tape. These benefits are applicable to as many as 60 missile stage units

per year. Natural follow-on efforts could be directed to the first, second, and fourth stages. Of the 34 improvement subprocesses identified during the activity week, 18 were implemented in real-time, 5 more were accomplished in the next two weeks, and 7 were implemented in the following month.

TARGET PROGRESS REPORT
MSPF TO LAUNCH FACILITY

27 Nov - 1 Dec	Monday	Tuesday	Wednesday	Thursday	Friday	Target
Set-Up Time (Hrs)	.85	.85	.85	0	Run New Process	.5
Cycle Time (Hrs) (includes set up)	9.5	5.3	4.3	2.4	Run New Process	4.8 3.5*
Man-hours	39.5	26.3	22.5	12.0	Run New Process	17.5
Walking Distance (miles)	.57		.44	.21	Run New Process	
Processing Steps	47		34	21	Run New Process	33

NOTE: *Stretch Goal Set

TARGET PROGRESS REPORT
MSPF TO DEPOT

27 Nov - 1 Dec	Monday	Tuesday	Wednesday	Thursday	Friday	Target
Set-Up Time (Hrs)	.75	.75	Not Observed	.6	Run New Process	.5 hrs
Cycle Time (Hrs) (includes set up)	6.75	3.75	Not Observed	2.5	Run New Process	3.4 hrs 3.0 hrs*
Man-hours	29	18.75	Not Observed	12.5	Run New Process	20 15 hrs*
Walking Distance (miles)	.56	Not Observed	Not Observed	.20	Run New Process	.28
Processing Steps	36	Not Observed	Not Observed	18	Run New Process	25 20*

NOTE: *Stretch Goal Set

Conclusion

Although national security is difficult to quantify, the benefits of the Peacekeeper reengineering effort are measurable. Reduction in stage processing cycle time produces increased combat readiness and emergency wartime surge capability. These advantages can be replicated at the Ogden depot and at Vandenberg AFB. Perhaps the most profound improvements relate to the human side of the organization. Decreased cycle time, transportation, and process steps means more time available for personnel development and opportunities which impact job satisfaction. These include job skills training, career broadening, site visits to Ogden and Vandenberg for benchmarking initiatives and sharing of best practices, and more time for streamlining other processes related to Peacekeeper missile processing. Follow-on reengineering initiatives being considered include space launch scheduling and Global Positioning System (GPS) modernization.

Reengineering Our Organizations: A Leadership Challenge

John Micalizzi
USAF Academy

Abstract

Reengineering organizations involves reconstructing key business processes fragmented by the trend of over-specializing jobs during our country's Industrial Revolution. As workers were forced to maintain strict performance standards for piecemeal tasks, organizations correspondingly developed layers of management to direct and control their behavior. The result was an increased alienation between workers and the results of their efforts. Reengineering represents a systematic approach to reverse this trend. Cross-functional process are put back together and managed by empowered teams. Leaders orchestrate the shift from task-based to process-based labor. They formalize the reengineering approach through a comprehensive strategic planning process and act as catalysts to initiate and sustain the effort. Despite what many theorists say, any size work unit can benefit from reengineering if leaders personalize the commitment to change. These small-scale initiatives may ignite larger efforts to incorporate reengineering thinking into an organization's overall planning process.

Corporate America has a long history of embracing the latest business trend promising to create a motivated and productive workforce while increasing the numbers at the bottom line. Management by Objectives, The One Minute Manager, Quality Circles, Zero Defects, and Total Quality Management, among others, have all been popularized at some time in the literature as the next, great philosophy to "save the company." Each has vowed to produce some radical rethinking of the types of jobs, processes, and management styles currently employed in many modern organizations. Today, businesses are struggling with ideas such as reinvention, reorganization, restructuring, reeducation and now, reengineering, which are being used to describe, and sometimes hide, the current trend of downsizing organizations and eliminating layers. Skeptical workers and managers at all levels are concerned that increasing efficiency may mean improving themselves out of a job. It's no wonder that the credibility of new, "revolutionary" approaches is being questioned at every turn.

Introduced by Hammer and Champy (1993) in their book, Reengineering the Corporation, the concept of reengineering was defined as "the fundamental rethinking and radical redesign of business processes to achieve dramatic improvements in critical, contemporary measures of performance, such as cost, quality, service and speed." (p. 32) Successful applications of this reengineering approach were reported at such diverse corporations as IBM, Ford, Hallmark, Taco Bell and Kodak. The centerpiece of this method is the strong commitment to a "clean sheet" approach, whereby broken processes are not continuously improved in small increments over

time, but rather, discarded and replaced by completely new processes. Reengineering's claim that "starting over" is the only way to truly improve organizations may be frustrating to some leaders and mangers who are still smarting from disappointing restarts with other, paradigm-shifting programs.

The principle purpose of this paper is to advance the idea that there is room for an evolutionary, iterative strategy for reengineering that incorporates the key ideas from Hammer and Champy (1993) while giving workers and management credit for the good things they are already doing. Insisting that everything must reset to zero before reengineering takes place may be unrealistic and unnecessary. Three objectives follow from this purpose: First, to show that modern reengineering is really the continuation of a relatively recent trend of recreating work processes from the fragments of over-specialization. Second, to outline the crucial role of leadership in initiating and sustaining reengineering efforts in their organizations. And finally, to propose a strategy for leaders to implement a reengineering approach in any work unit, regardless of size or complexity, that would shift the focus of labor from tasks to processes.

Leaders at all levels are striving to construct a context for their organizational priorities. What should be the focus of their improvement efforts? Greater profits? Higher performance? Lower costs? Reduced cycle times? Whatever outcomes are selected, leaders must link them to the work that subordinates are doing on a day-to-day basis. Unfortunately, work in general has traditionally been fragmented into minute tasks and delegated to individuals who sometimes lacked the training, resources, authority and, therefore, the ownership for improving their jobs. This trend toward work fragmentation has its roots deep in the early years of our industrial revolution.

The Rise Of Bureaucracy

As the United States shifted from a nation of craftsmen and artisans to one of industrial dominance in the early part of the twentieth century, a strict division of labor began to emerged between workers and the new "management" level. Workers with little education and experience provided the vast energy source needed to fuel the industrial machine. Engineers and businessmen at the top of the pyramid performed planning, coordinating and organizing functions while workers at the bottom performed the mindless work of automatons, capable of performing only the most simple, highly structured and repeatable pieces of work. Control and conformity were viewed as essential to satisfy the consumer's insatiable appetite for the latest products which placed a higher priority on availability rather than quality.

Dr. Frederick Taylor was instrumental in revolutionizing industrial work through careful time-motion studies which claimed to find the one most efficient, "best" way to perform a given task. Even the most complex job was believed to be "reducible" to piecemeal steps that virtually any well-trained individual could perform at a consistently high level for long periods of time. But the need to maintain strict, uncompromising consistency over one's work behavior drove business leaders to create a new level of supervision heretofore unknown in the history of American labor. This new, "middle-level" of management between the planners and the doers would be responsible for training, controlling, evaluating and correcting worker behavior that deviated from the

standard way. In short, the mission of middle management was to enforce the plan from above. Workers were not expected to be creative or to even care about improving their work. Someone else would do that for them. Their only job was to perform the work as directed.

The Impact of Work Fragmentation

The result of this trend was three-fold. First, it created a workforce alienated from responsibility, authority and accountability for improving their work. Innovation was unthinkable and actually discouraged since it represented a deviation from the norm. Workers were not being paid to think, but to perform their jobs within prescribed limits. Second, the fragmentation of work spread to the organization itself. Functional responsibilities were "stovepiped" into separate departments. Planning, for example, now became the responsibility of a group of professional planners whose job it was to plan for the entire organization. Furthermore, when other new responsibilities were added to the organization, whole new departments appeared defined by their unique functional expertise. As a result, organizations began to expand both vertically and horizontally as various forms of departmentalization evolved in search of the best way to group work tasks.

The most disturbing result of this trend was that no single individual or group was managing the cross functional work-flow across departments. To make matters worse, this horizontal workflow usually resulted in the products and services delivered to external customers. Since no single individual or group felt responsibility for the entire workflow that produced a product, the "mindset" of the organization became focused on the small pieces of work defined by their functional responsibilities. Planning, inspection and improvement were now someone else's job. This fracturing of work processes produced the simplicity and repeatability of Ford's assembly line, but at a cost of an overly complex management structure that increasingly distanced workers from ownership of the ultimate product of their labor.

Breaking The Paradigm Of Specialization

Since that time, organizations have been gradually picking up the pieces of their key business processes. Two major developments have hastened this reconstruction. First, people in general are coming to work with a dramatically different set of expectations about their jobs. More years of formal education, higher economic status, a greater number of choices available, among other factors, have created a workforce that comes to the job expecting to make a contribution, not just carry out orders. Workers want their work to mean something, they want to be appreciated by bosses and co-workers, and they want a say in improving their jobs. There's no going back to "checking your brain at the door" when you enter the plant. Second, outside competition to provide goods and services has created higher customer expectations about the products they buy. Quality is now more important than availability. Customers have more choices then ever to satisfy their needs. As a result, organizations are being forced to determine customer needs and then deliver the products and services to satisfy them. Piecemeal work performed by unmotivated and disinterested workers in compartmentalized jobs will no longer serve or keep customers very long.

A Leadership Strategy For Reengineering Organizations

Reengineering involves putting back together the processes that Taylor, Ford and other early industrial scientists tore apart in the name of efficiency. In these reengineering efforts, senior leaders are thrust into the role of architects, designing process-based systems where empowered teams lead improvements in their work. Traditional managers become coaches and mentors, guiding their people to assume greater responsibility and become leaders in their own right. Both now play vital roles in reengineering the organization by formulating, deploying and implementing an overall planning process for charting the future direction of the company.

The potential power of reengineering can be more fully realized when it is integrated into an overarching, systematic, strategic planning process. Leaders at all levels must provide the energy to initiate and sustain the planning effort. They set the tone, establish the values, articulate the core mission, and help develop a shared vision of the organization's future. The leader's commitment to planning insures that decisions concerning process structure, definition, analysis and measurement are made based on a thorough review of the overall mission, vision, and goals of the organization. Work processes become linked and aligned with management information systems as well as the priorities of the organization. Contributions to process improvement are recognized and rewarded. Loyalties to functional areas are replaced by commitment to customer satisfaction and organizational improvement.

Leaders also become the catalysts for changing preconceived notions of how work is done in organizations. They challenge workers to abandon the task mindset and focus on the bigger picture of delivering products and services to satisfy customers. Functional walls are broken down as key work processes are now permitted to easily flow through the stovepipes on their way to customers. Leaders promote a culture of trust and respect as they act to reassure workers that reengineering work processes will help them better serve their customers as well as advance their own personal growth and leadership development. As the most visible agents of organizational change, leaders must serve as champions of this new vision.

Finally, leaders take on the crucial role of enabling and empowering workers to successfully manage these larger processes. As guides and coaches, leaders set the standards for a caring mentorship that pervades the entire organization. Resources are provided to convert the vision into reality. Traditionally trained as expert problem solvers in their functional areas, managers and workers must be reeducated to assume responsibility for complex, cross functional, processes. As competence and maturity grow, workers are granted more autonomy to make decisions and innovate changes. Dealing with ill-defined issues in complex environments now becomes the norm. And, as the manacles of control and bureaucracy are removed, workers become free to take reasonable risks to improve the organization. Leaders play an important role in building this culture of trust by rewarding workers who improve their processes and not punishing those sincere efforts if they fail.

Where Do I Begin In My Organization?

It is unlikely that many large organizations will choose to implement a comprehensive reengineering of their key business processes from the ground up. Especially if marketed as "radical" and "dramatic," a formal reengineering program will probably scare off more potential implementers than it will attract. Does reengineering have practical relevance for today's leader?

The principles of reengineering described in this paper can be effectively employed in any size organization, at any level, and within any work setting. It begins with committed individuals and groups who decide for themselves that they will benefit by reconstructing processes. Formal and informal leaders must define the group's sphere of influence. Since reengineering may not be instituted organization-wide, boundaries with other functional areas should be identified and tested. Other workers may also be interested in linking up with their processes. Leaders must focus on the critical mass of people sharing a common purpose. Representing this group, the leader must communicate the mission and operationalize the vision of the larger organization. A shared strategy for rebuilding work processes and enabling and empowering group members must be developed, deployed and linked to any higher-level planning effort going on.

Planning to reengineer begins with identifying customers that use or need the products and services that your group is responsible for delivering. Providing your customers with a single point of contact as you seek to understand their needs is probably one of the most valuable services you can offer. Rebuilding processes with the end result in mind will help to insure that the flow of work directly contributes to satisfying customers. Leaders in concert with workers must establish authority and responsibility to manage these different processes. Comparing the new process to its "as is" version (if there is one) will generate opportunities for improvement. Process experts now develop appropriate indicators to display the "health" of the process. Data are collected and adjustments made based on the evidence. Finally, process improvements are standardized and institutionalized and the next review cycle begins. In this way, continuous process improvement methods are integrated into a reengineering strategy.

Most small-scale reengineering efforts can be accomplished without having to ask for specific permission from higher authority. The key is to focus on activities you control within your sphere of influence. Obviously, the higher the reengineering effort is initiated, the more pervasive will be its impact. But, regardless of the level of implementation, everyone needs to do what they can, do it now, and do it better than anyone else. It would be unreasonable to expect others to immediately embrace any change in their work routine. Change can be confusing and frightening. Leaders set the context and the framework for dealing with change. Under their caring guidance, workers are encouraged to take that 10% stretch to accomplish more than they thought possible. Acting empowered and taking the initiative may bring further responsibility and authority from leaders who value this behavior. Reasonable risk-taking becomes a prelude to innovation as you continually test the limits of tolerance for new ideas. The stage is set for leaders to turn barriers into opportunities for improving both the work and work life in modern organizations.

References

Hammer, M., and Champy, J. (1993). Reengineering the Corporation. New York: HarperCollins.

Executive Coaching: How to Achieve Long-Term Leadership Behavior Change

Gordon J. Curphy, Ph.D.
Vonda K. Mills, Ph.D.
Personnel Decisions International

Abstract

Over 3,000 managers and executives have gone through one of Personnel Decisions International's customized coaching programs. This panel presentation will discuss an actual case study of an executive coaching client. The panel will describe the situation leading up to the coaching intervention, the coaching assessment process, several behavioral change modules and techniques, and the results of the coaching program to date.

Executive and management coaching programs are becoming more popular as we approach the 21st Century. This rising popularity is primarily due to two factors. First, organizations are beginning to realize that training interventions alone are often not enough to cause permanent behavioral change. Too many managers go through a leadership training program, return to work with intentions of changing their attitudes and behaviors, but three months later revert back to their previous work habits. Second, as people move to the top levels in organizations it often becomes necessary to customize developmental interventions to meet individual needs.

This panel presentation will provide a broad overview of what executive coaching is (and is not). We will then describe a theoretical model and some of the research on coaching effectiveness. The last part of the panel presentation will consist of a case study of an executive coaching client. This case study will include a discussion of the organizational context and the situation leading up to the decision to use a coaching intervention. We will then describe the personality, mental abilities, work simulation, multirater feedback, and interview results used to determine the content of the coaching program. We will conclude the presentation with a description of several of the coaching modules and progress to date. As this is a panel presentation, audience participation will be highly encouraged.

Panel Proposal for
The 15th Biennial Applied Behavioral Sciences Symposium

Bridging the Gap Between Leadership Outcome Development and Their Assessment: Easier Said than Done.

LCDR R.R. Albright, U.S. Coast Guard
CDR P.T. Kelly, U.S. Coast Guard

This panel session will focus on the difficulties encountered when institutions begin to develop assessment schemes designed to measure Leadership Outcomes that have been participatively created and agreed upon by faculty and military professionals. To begin discussion the process undertaken by the Coast Guard Academy faculty to create Leadership Outcomes will be described. The outcomes the Academy desires in its graduates will be presented and the initial efforts the Academy has made toward the assessment of these outcomes will be discussed. The results of a survey conducted to measure the extent to which Coast Guard Academy seniors have mastered the desired leadership outcomes will be presented. The survey had been constructed directly from the descriptive statements that comprise the institution's desired Leadership Outcomes. Factor analysis results infer that a number of the outcomes may lend themselves to assessment via survey methodology. However, several of the outcomes failed to empirically load as they had been conceptually created. These early results infer that the descriptions of several of the outcomes need to be re-examined for internal consistency and overlap.

CGA Leadership - Outcomes

These outcomes make the following assumptions:

Leadership development is a core function of the Coast Guard Academy, and should be a central part of every program at the Academy. Academic, military, athletic, and social programs should all directly contribute to leader development. We desire that our graduates receive the education and training that will prepare them for a career of service as future leaders of the Coast Guard. Each program, division, and course should be examined in light of these Leadership Outcomes to further cadet leader development.

Leadership outcomes will be assessed to the largest extent possible. Leadership outcomes include commitment to an underlying set of values and virtues, and thus may defy easy quantification or measurement.

USCGA graduates shall:

1. Demonstrate understanding and usage of leadership theories when serving in a leadership role.

2. Demonstrate moral and ethical judgment.

3. Demonstrate the ability to direct and develop others.

4. Demonstrate facility in functioning up, down, and across a chain of command.

5. Demonstrate the ability to function as an effective team member.

6. Demonstrate respect for all persons one interacts with as part of one's role and areas of responsibility.

7. Demonstrate professional decision-making ability.

8. Demonstrate professional communication ability.

9. Demonstrate an ability to self assess their leadership ability.

10. Describe a personal framework of leadership that integrates the Core Values of the Coast Guard.

2. Demonstrate moral and ethical judgment.

Discussion: Coast Guard Academy graduates must have the ability to make moral and ethical choices that can be defended. Every cadet arrives at the Academy with a system of personal values. These values may relate to honor, ethics, leadership, human relations, and general personal conduct. As officers, our graduates are expected to adhere to a certain level of behavior regarding their professional and personal conduct. This results in the requirement for the Academy to educate and train them in areas of personal and organizational values. Our graduates must be able to recognize ethical considerations associated with their actions and behavior, identify alternatives in difficult ethical choices, systematically analyze the conflicting considerations supporting different alternatives, and formulate, defend, and effectively carry out a course of action that takes into account this ethical complexity.

Implementation: The commitment of cadets to being moral and ethical people is fundamental to the Academy's mission. The cadet experience is a daily validation of moral and ethical judgments. When, how, and why are incorporated into the ethical framework of cadets developing into officers. The immediate approach of knowing when to do the "right thing" is judged through academic, military and athletic endeavors. A good example is understanding the Honor Concept through the program presented during 4/C summer. The core course selection in Morals and Ethics requires the examination of moral and ethical theory, while the Ethics Forum and Honor Week Sessions reinforce annually the moral and ethical code of behavior cadets are expected to know and maintain. Finally, the application of moral and ethical principles is reinforced in upper class leadership courses.

Assessment: The measuring of the cadet's moral and ethical judgment is an ongoing subjective appreciation of how the cadet is evaluated in his or her company. Peer judging is important in the perception of how the cadet is able to function with other cadets in leadership roles. An example here is Cadre summer performance in making quality, ethical choices with 4/C cadets. Feedback (positive, negative, internal, external) is based on he evaluative process set in place. Performance reports seem to be the driving force in understanding the moral concept; a restraining force is the obvious misevaluation of the cadet. Another primary tool of assessing is the observed "pressure decision situation" associated in athletics. When the cadet, as an athlete, is in a position to make decisions in a quality fashion under the observance of fellow cadets, superiors, family and friends, the moral fiber of the human being is exposed in a vulnerable position. The objective of making correct choices in times of duress or stress places cadets into familiarity zones of ethical considerations as they develop into career officers. Finally, assessment of Academic Division Outcome # 10 provides valuable information that can be used to evaluate performance relating to this Outcome.

Level 1: Cadets are able to recognize and articulate their values and those of the Coast Guard.

Level 2: Cadets are able to employ basic ethical concepts and reasoning when presented with typical cases relating to the Academy and Coast Guard. These include cases involving honor issues, ethical conduct, inappropriate relationships, and professional conduct. Cadets are able to

identify ethical issues and choices, to evaluate critically alternative ethical courses of action, and to defend a selected course of action.

Level 3: Graduates are able to integrate the values developed at the Academy into their own conduct. This includes moral reasoning, judgment, communications, and interpersonal skills to carry out duties and responsibilities. Furthermore, graduates are able to develop moral and ethical courses of action for relatively complex ethical challenges facing leaders in general, and Coast Guard officers in particular.

The Fundamental Role of Leadership: Developing Followers Into Partners

William E. Rosenbach, Ph.D.
Thane S. Pittman, Ph.D.
Gettysburg College
Earl H. Potter III, Ph.D.
Cornell University

An important thread within the new paradigm of transformational leadership is the concept of leaders' transformation of followers into leaders. One version of this can be found in the recent work of Manz and Sims (1993) on superleaders. They argue that a central role of leadership for the new millennium will be the development of followers who can exert self-leadership. Still another view, put forth by Rost (1993) is that leadership is best understood not in terms of personal characteristics of the leaders or their actions but in terms of the dynamic interaction between leaders and followers. At some point, Rost argues, the distinction between leaders and followers becomes both artificial and, ultimately, meaningless.

We wish to present a related but somewhat divergent approach, one which maintains what we see as an essential distinction between leadership and followership but which also recognizes that neither has meaning except in the context of the other. We propose to share our recent work, both conceptual and empirical, in this area. We will present the Performance and Relationship Questionnaire which we have developed in consulting work in the U.S. and internationally which allows leaders and followers to assess the nature of follower styles.

New Horizons in Leadership:
Creating Organizational Rainbows

The Connectivity Between Leadership, Creativity, Innovation, and Change

Summary:

The purpose of the panel entitled "New Horizons in Leadership: CreatingOrganizational Rainbows," is to examine leadership focused on creativity, innovation and change today. We have put together a panel of experts, notintimidated by the challenge of creativity, who are helping train leaders forthe future. Creating the rainbow is the new "formula" for corporate, government, and personal success. Is creativity becoming the new bottom line? How can leaders help people recreate or rediscover themselves? Can imagination bemanaged? What broad-based knowledge and experiences are key factors for innovation? Where is it happening now? What makes us think that creativityand innovation are answers? What stops us from inspiring new directions, new ideas, new dreams? What causes us to be so stuck in the present? Whatwould happen if this creative leadership territory took the place of quarterlyfinancial results?...What will happen if it doesn't?

The panel discussion will bring out a clearer understanding of how leaders can enhance and apply creativity. Panel members and the audience will learn moreabout the connectivity between leadership, creativity, innovation, and change.

Colorado Tech Panel Members:

Dr. Frank Prochaska
Professor of Management
Colorado Tech
Consultant

Dr. Vicki Strunk
Professor of Management
Colorado Tech

Dr. Bill Wallisch
Doctorate of Management (Candidate)
Consultant

Don Marble
Doctorate of Management (Candidate)
Associate Professor
Colorado Tech

Jerry Reinsma
Doctorate of Management (Candidate)
UTMC Senior Manager

Betty Rosengren
Doctorate of Management (Candidate)
MCI Senior Manager

Imagination

by Frank J. Prochaska, Ph.D.

The only resource organizations have is imagination. Emboxment stops imagination. It is vital that the leader knows how to design a structure and environment which allows people, including the leader, to grow, and imagine. This takes courage, creativity, and, above all, self-mastery.

Dr. Prochaska designs and teaches creativity and leadership courses at Colorado Technical University. He is a mentor in the Doctoral program.

Creativity and Innovation in Change Leadership

by Vicki L. Strunk, Ph.D.

Creativity and innovation in leadership is most evident in organizational change, and change leadership is becoming a core competency for the twenty-first century. The companies that survive in the coming decades will be those that are able to respond proactively to changing environmental conditions. This simple premise is easy to understand but difficult to put into practice. Experience and broad-based knowledge are probably the key factors in innovative leadership: The more experience a leader has, the more confidence he is likely to have in his ability to solve problems and implement changes, and the more risk he is willing to take. The leader who constantly combs the total environment, not just his industry, for new ideas, will more likely find innovative solutions and adapt those solutions to his organization's needs. Thus, creativity and innovation are not necessarily defined solely as the uniqueness of a solution, but also the originality of its adaptation.

Dr. Vicki Strunk is Professor of Management at Colorado Technical University and has been teaching there since 1989. She is also academic advisor for the Department of Management. Her doctoral research focused on transformational change in organizations and the leadership factor in that change. She is currently serving on several dissertation committees.

Creativity: The New Bottom Line

by Bill Wallisch

More and More, creativity and innovation are the key attributes a corporation must have in order to make it to Fortune's list of "Most Admired" companies. In the past, quarterly financial results were everything. Now there's a growing understanding that corporations can't live by numbers alone. From personal experience as a consultant to Fortune 100 companies I can tell you that "releasing the power" of creativity and innovation is the number-one leadership focus, from the board room to the plant floor. CEOs are doing everything they can to stimulate ideas, create a climate of creativity, and help generate a new sense of corporate "dreaming." Leaders who can't inspire ideas are being forced to strap on their golden parachutes and bail out. The drive is on for new directions, new ideas, new products, new markets, and new processes. Bottom line: creativity. How to survive as a corporate leader? Get people thinking. Chief frustration? "How do you make that happen?" I'd like to share some of the successes and failures I've seen as top corporate and government executives strike out to explore new -- often unfamiliar and treacherous -- creative leadership territory.

Dr. Bill Wallisch is a retired Air Force Lieutenant Colonel, tenured Air Force Academy professor, and administrator. He is the creator of the Academy's "Blue Tube" and also served as both director of public affairs and executive to the Superintendent.

Re-creation, not wreck-reation

by Don Marble

It isn't "wreck-reation" at all: it is RE-CREation. Re-creation of mental models and Re-creation of self places effective leaders/managers apart from the rest. It is the leader in the role of designer (in the sense of the architect) of the organizational structure that provides the environmentally friendly framework for shared vision and a domain for communal motivation. An organization is the sum of the products of the various interacting groups. The job of the designer is to assure that the signs of these products are all positive so that the overall effect of the group's interactions are additive.

Don Marble was a Senior Systems Management Executive with Litton in large military tactical data systems. His experience is in program and project management advanced systems development, engineering systems test, and quality assurance.

Vaulting Barriers to Creativity

by Jerry Reinsma

With the globalization of the marketplace and manufacturing capabilities, traditional competitive advantages such as technology, capital equipment, and financial resources are starting to level. In the future, the most significant source of a sustainable competitive advantage will lie in the creativity and innovative abilities of the human resources of the enterprise. Major barriers to developing the creative potential of the workforce include fear, conformity pressures within team environments and the inability to deal constructively with the tension-resolution dynamics inherent in the creative process. Until these obstacles are dealt with, and the organization proactively endorses and values creativity, the financial impact to the bottom line will be minimized.

Jerry Reinsma is the Director of Operations and Quality Assurance at United Technologies Microelectronics Center in Colorado Springs.

Self-Discovery

by Betty Rosengren

Self-discovery is a wonderful thing for leaders. Dr. Prochaska's paradigm shifts in the beginning of the Creative Leadership course made me very uncomfortable. Later, I found them to be the neatest thing for me. Senior leaders have to rediscover themselves. This new-found creativity continually opens up new doors to the future.

Betty Rosengren is a Senior Manager with MCI international operations. She has a Masters degree in Nursing and was a nurse for many years, including work at the Mayo Clinic.

Personality and Leadership: What do we know about Selection, Training, and Development?

Gordon J. Curphy, Ph.D.
Personnel Decisions International
Captain Kevin D. Osten, M.S.
Colorado State University
Lieutenant Jeffrey Voetberg, M.S.
Occupational Measurement Squadron

Abstract

Personality has enjoyed a resurgence in popularity among industrial/organizational psychologists in recent years. This panel will discuss what is currently known about the use of personality assessment in both military and civilian settings.

Over the past thirty years personality has waxed and waned in popularity among organizational psychologists. Thirty years ago the popularity of personality assessment was at its nadir, with Guion and Gottier (1965) stating that there appeared to be no compelling evidence to support the use of personality assessment in selection or training. Over the past five years, however, a number of researchers have published articles showing strong support for the use of personality assessment in selection across a wide variety of jobs (Barrick & Mount, 1991; Curphy & Nilsen, 1995; Hogan, Curphy, & Hogan, 1994; and Tett, Jackson & Rothstein, 1991). This panel will discuss what personality is, how it can be measured, why it has recently become more popular, and how it can be successfully used in selection, training, team building, and development settings.

In terms of specific contributions, Dr. Curphy will discuss his research and practical experience working with over ten different personality assessment tools in a variety of military and civilian settings. He is also the co-author of an American Psychologist article summarizing eighty years of personality and leadership research. Captain Osten will describe his research using personality assessment as a part of a training evaluation study involving ROTC instructors. Lieutenant Voetberg will discuss his comparative research examining the relationships between personality traits and ROTC student versus Air Force Academy cadet performance.

The Leadership Development Survey in a Reserve Officer Training Corps Setting[1]

1st Lieutenant Jeffrey W. Voetberg, M.S.
Air Force Occupational Measurement Squadron

Abstract

The Leadership Development Survey, a personality-based 360 degree feedback tool, was administered to members of the Reserve Officer's Training Corps. Three observers also rated each target individual. Factor analysis did not support the structure of the LDS, but agreed with previous research with the instrument. Contrary to expectation, LDS scores were not predictive of performance. Implications are discussed.

With the failure rate among senior leaders in this country hovering around fifty percent (DeVries, 1992), it has never been more important to provide the feedback helpful for success. One area in which feedback is helpful is leadership. The benefits of leadership feedback have been demonstrated in theory (Hughes, Ginnett, & Curphy, 1993) and in practice (Atwater, Roush, & Fischthal, 1995). The Leadership Development Survey (LDS) was designed to help leaders by measuring the personality traits associated with leadership and providing feedback on those traits.

The Leadership Development Survey

Based on the recommendations of Nilsen and Campbell (1993), the LDS is based on both self and other's observations. No constraints, other than a familiarity with the target, are placed on the nature of the observers, thus it is possible to access inputs from subordinates, peers, and supervisors. The two versions of the LDS, the self form (LDS-S) and the observer form (LDS-O), contain slightly different demographic but conceptually identical substantive items. All substantive items are rated on a six-point scale, where one is "strongly disagree," and six is "strongly agree."

The 66 items are based on the Big Five traits of Extraversion, Conscientiousness, Neuroticism, and Agreeableness. These four traits have been repeatedly found to be positively correlated with leadership behavior ratings (See Hughes, Ginnett, and Curphy (1993) for a summary). The other Big Five trait, Openness to Experience, has not been found to be related with ratings of leadership, and was therefore not included in the LDS. The four major scales were further divided into 11 traits or subscales.

The items in the survey are based on personality traits, but are behavioral in nature. It is possible that perceptions of an actor's behavior will change according to the status of the observer. The individual's actions may be viewed in one way by her or his subordinates, and

[1] Portions of this paper were submitted in partial fulfillment of the requirements for the degree of Master of Science from the University of Illinois. The views expressed in this paper are the author's own and do not necessarily reflect those of the Air Force Occupational Measurement Squadron, the Air Force, or the Department of Defense.

another way by peers or supervisors. The LDS is designed to measure all perceptions, and report the results back to the individual.

After its initial administration in 1993, the LDS has been periodically administered to upperclassmen at the Air Force Academy (AFA) as part of a leadership course. Unfortunately, there are no data available other than from the initial administration, which consisted of 257 juniors and seniors as targets, and 1,008 observers.

Personality and Performance

Several major reviews and empirical studies have demonstrated the link between personality traits and performance. Stogdill (1974) found several traits which were related to rated leadership performance. We now organize the traits under the Big Five taxonomy as Extraversion, Dependability, Agreeableness, and Adjustment.

Perhaps more importantly, the relationship between personality traits and effective team performance is well supported by theory and data. Hogan, Curphy, and Hogan (1994) explain how more extraverted, dependable, agreeable, and well adjusted leaders are better able to build and maintain more effective teams.

The current study uses the LDS with members of the Air Force, Army, and Navy Reserve Officer Training Corps (ROTC). The structure of the LDS will be examined using exploratory factor analysis. It will be determined how well the four-factor version of the LDS is reflected in the data. It is hypothesized that this sample will support either the rational four-factor structure or the three-factor model which was found in the initial administration.

Additional analyses will examine the correlation between LDS scores and a measure of performance. It is hypothesized that the Extraversion and Dependability scales will be significantly correlated with rated performance.

Method

Measures

The LDS was used to measure the personality traits of each individual. Demographic variables were assessed using nine items at the beginning of each survey. Observers were asked about their relationship to the ratee, while target individuals were asked their current class standing, grade point average, and current military position. At the completion of the survey, respondents were asked questions regarding the rater's subjective accuracy (for observers) or the pressure to present themselves "in the best possible light" (for the self form).

Each target individual's order of merit (OM) was obtained through self-report. ROTC members are rank-ordered within class by the officer in charge of training. This rank-ordering takes into account grade point average, performance in military education classes, and a subjective measure of leadership effectiveness.

Subjects

All subjects were members of ROTC at the University of Illinois. Individuals without both subordinates and superiors were not asked to participate. Subjects were told that the results were for developmental feedback only and would not be made available to their supervisors. Participation was strictly voluntary.

A total of 101 subjects were approached regarding the study: seventy-seven returned usable responses. Fifteen were Air Force (83% response rate), thirty were Army (79%), and thirty-two were Navy (71%). Of these seventy-seven, 68 (88%) were male and ninety-one percent were Caucasian. Seventy-three percent of the sample indicated their age to be between twenty and twenty-two.

These 77 individuals distributed the observer forms to 231 individuals. 223 returned usable surveys, a 97% response rate. Observers' demographics mirrored the targets'.

Procedure

Subjects were responsible for identifying the three observers. The observers were to be a military subordinate, superior, and a peer. Peers included anyone in the same class year or with equal military status. Observers returned their forms to the officer in charge of training for each service branch to insure the target did not have access to an individual observer's ratings. Each target received her or his scores and an interpretive guide. The scores reported on the feedback sheet were standardized to a mean of 50, with a standard deviation of 10.

Results

As this was the first administration of the LDS with an ROTC sample, it was necessary to establish the internal consistency and reliability of the LDS in this sample. This was done by examining the corrected item-to-scale correlations and Cronbach coefficient alphas. Table 1 shows the Cronbach alphas for each scale and subscale. For comparison, the alphas from the initial administration are included in Table 1. The item-to-scale correlations ran from .21 to .80. The LDS appears to be a reliable instrument, across samples, situations, and service branches.

Since past analysis did not support the rational structure of the LDS, an exploratory factor analysis was performed on the results of this administration. The initial principal factor analysis retained three factors for both the self and the observer versions. These factors were then subjected to both orthogonal and oblique rotations. The varimax orthogonal rotation produced the most interpretable results. The factor loadings can be found in Table 2. The consistency between this solution and that from the AFA administration is excellent. Both produced a three-factor solution, and the components of the factors were nearly identical. For comparison, the AFA sample factor loadings for the observer forms are included in Table 2. A word of caution is needed here however:- The target sample size is smaller than recommended by Ford, MacCallum, and Tait (1986). The dual loading of Achievement-Orientation and Conservatism may be noise resulting from the small sample.

Table 1
Cronbach's Coefficient Alphas for All Scales and Subscales

	Current Study		AFA	
Scale and Subscale	Self Form	Observer Form	Self Form	Observer Form
Extraversion	.82	.91	.88	.86
Dominance	.82	.91	.81	.80
Sociability	.83	.91	.85	.85
Dependability	.82	.90	.86	.92
Achievement Orientation	.82	.90	.81	.86
Conservatism	.84	.91	.78	.83
Organization	.85	.91	.74	.77
Credibility	.81	.90	.72	.85
Agreeableness	.81	.90	.81	.89
Friendliness	.82	.90	.65	.78
Empathy	.83	.90	.68	.81
Likeability	.82	.90	.80	.87
Adjustment	.81	.90	.85	.85
Emotional Stability	.83	.90	.77	.80
Self-Acceptance	.81	.90	.80	.75

Table 2
Factor Analysis Results for the LDS Secondary Scales

	Factor Loadings ROTC Sample Self Form			Observer Form			AFA Sample Observer Form		
Secondary Scale	I	II	III	I	II	III	I	II	III
Sociability	.76			.77					.80
Likeability	.75			.71					.61
Dominance	.69			.70					.57
Self-Acceptance		.79			.77		.79		
Friendliness		.70			.80		.87		
Empathy		.50			.73			.61	
Emotional Stability		.49			.66		.66		
Achievement-Orientation	.47		.67			.83		.88	
Credibility			.62			.74		.60	
Organization			.50			.77		.71	
Conservatism		.31	.31			.71		.67	-.35

Spearman's rho correlations for ranked data were computed between the target individual's OM within her or his class and each of the scales and subscales. Results can be found in Table 3. There were no substantial differences in the correlations between the three service branches, so the correlations were averaged across the branches, using sample size for weights. Contrary to expectation, only one scale, the self version of achievement orientation, was significantly correlated with OM, although most of the correlations were in the predicted direction. Credibility on the Dependability scale showed a fairly strong correlation, as did dominance, though they were not significant.

Table 3
Spearman's Rho Correlations between Order of Merit and LDS Scores

	ROTC Sample		AFA Sample	
Scale and Subscale	-S	-O	-S	-O
Extraversion	.13	.04	.03	.10
Dominance	.23	.11	.19*	.30*
Sociability	.08	-.02	-.07	-.07
Dependability	.19	.06	.47*	.50*
Achievement Orientation	.37*	.09	.38*	.47*
Conservatism	-.06	.02	.44*	.47*
Organization	.06	.05	.24*	.37*
Credibility	.22	.05	.27*	.37*
Agreeableness	.1	-.01	.01	.10
Friendliness	.01	-.01	-.02	.05
Empathy	.09	-.04	.03	.14
Likeability	.09	.00	.01	.06
Adjustment	.08	-.02	.00	.05
Emotional Stability	.00	.02	.00	.03
Self-Acceptance	.11	-.03	.00	.07

* indicates significance at the .05 level

Discussion

One major problem with the LDS is the fact that exploratory factor analysis has now twice failed to support the rational structure of the LDS. The fact that the current results are nearly identical to the original findings suggests that the true structure is composed of three factors, not four. Factor I is the original Extraversion scale plus the likeability subscale. The content and general feeling of these subscales are very similar. Dominance assesses assertiveness, self-confidence, and the tendency to control. Sociability is indicated by outgoingness and gregariousness. Likeability is similar to popularity, and may actually be the result of high dominance and sociability. Factor II is the combination of Agreeableness and Adjustment, minus the likeability subscale. The four subscales which comprise this factor deal with the individual's affect. The subscales measure how cheerful, optimistic, sensitive, calm, and comfortable the person is with him- or herself. In fact, the survey authors acknowledge the close relationship

between these four subscales in the technical manual (Curphy and Osten, 1993, p. 17). Factor III is the original Dependability scale, containing achievement orientation, conservatism, organization, and credibility. Future studies should attempt to confirm the three-factor model, attach meaning to the three factors, and determine their implications for leadership.

The lack of strong correlations with OM are somewhat disturbing. Past work with the LDS yielded strong correlations with the dominance subscale and the Dependability scale and subscales. While the general pattern was the same in this study, at least for the self reports, the correlations were less consistent and weaker. Contrary to expectation and past research (Harris & Schaubroeck, 1988), the self reports were, in general, better predictors than the observer scores.

One explanation is restriction in range in both variables. Thirty-eight cadets indicated that they were in the top ten of the OM. Only a few individuals used the lower rankings. In addition, the LDS items were almost without exception skewed toward the high end. Such restriction in range in both variables would tend to reduce the magnitude of the correlation coefficients.

Another explanation is the nature of the criterion measure. OM is a self-reported global measure, tapping several performance areas. Pulakos et al. (1988) demonstrated that a more homogeneous criterion will yield better predictor-criteria links.

The LDS is a relatively new instrument, and work still needs to be done. Future studies should attempt to better define the relationship between the LDS scales and performance. It would also be beneficial to observe the effects of feedback over time. Atwater et al. (1995) demonstrated the benefits of feedback, but it is not known how long the effects of a one-time intervention last.

References

Atwater, L., Roush, P., & Fischthal, A. (1995). The Influence of Upward Feedback On Self- and Follower Ratings of Leadership. Personnel Psychology, 48, 35-59.

Curphy, G.J., & Osten, K.D. (1993). Technical Manual for the Leadership Development Survey. (Tech. Rep. No. 93-14). Colorado Springs, CO: United States Air Force Academy.

DeVries, D.L. (1992). Executive Selection: Advances but No Progress. Issues & Observations, 12, 1-5.

Ford, J.K., MacCallum, R.C., & Tait, M. (1986). The Application of Exploratory Factor Analysis in Applied Psychology. Personnel Psychology, 39, 291-314.

Harris, M.M., & Schaubroeck, J. (1988). A Meta-Analysis of Self-Supervisor, Self-Peer, and Peer-Supervisor Ratings. Personnel Psychology, 41, 43-62.

Hogan, R., Curphy, G.J., & Hogan, J. (1994). What We Know About Leadership. American Psychologist, 49, 493-504.

Hughes, R.L., Ginnett, R.C., & Curphy, G.J. (1993). Leadership: Enhancing the Lessons of Experience. Homewood, IL: Irwin.

Nilsen, D., and Campbell, D. (1993). Self-Observer Rating Discrepancies: Once an Overrater, Always an Overrater? Human Resource Management Journal, 32, 256-281.

Pulakos, E.D., Borman, W.C., & Hough, L.M. (1988). Test Validation for Scientific Understanding. Personnel Psychology, 41, 703-716.

Stogdill, R.M. (1974). Handbook of Leadership. New York: Free Press.

Effects of Proximal and Distal Context Variables on Performance Appraisal Quality: A Model and Framework for Research

Kevin R. Murphy, Jeanette N. Cleveland, Christine Henle, Kim Morgan, Michael Orth
Department of Psychology, Colorado State University

Aharon Tziner
Universite de Montreal

Abstract

The conceptual model guiding a multi-year, multi-national effort to better understand the role of contextual factors in appraisal is described. This model leads to a number of predictions about the interrelations among context factors and the direct and indirect influence of these factors of raters likelihood of giving high or low ratings, willingness to discriminate good from poor performers, and willingness to discriminate among various aspects of job performance when completing ratings.

Recent models of the appraisal process (e.g. Cleveland & Murphy, 1992; Murphy & Cleveland, 1991, 1995) have suggested that a number of characteristics of both the appraisal system and the organizational context in which it resides are critical for understanding the rating process. So called "rater errors" and other shortcomings of appraisals in fact represent conscious decisions on the part of the rater to distort performance ratings in order to help attain personally or organizationally-valued goals. For example, a supervisor who wishes to motivate a particular subordinate might give that person higher ratings than his or her performance merits.

In order to understand the processes involved in deciding how to complete performance evaluations, it is necessary to consider both the rater's attitudes and beliefs that are immediately relevant to the task of evaluating performance (i.e. proximal influences) and his or her more general beliefs and attitudes toward the organization (i.e. distal variables). Empirical research on the roles of attitudes and beliefs about appraisal systems and the organizational contexts in which they reside in shaping rating behavior is just starting to emerge. For example, Tziner, Murphy and Cleveland (In press) report promising results in several initial studies.

The paper describes the conceptual model guiding a current multi-year, multi-national research program examining the relationship between several contextual variables and measures of rater behavior (e.g., likelihood of giving high or low ratings). One major theme of this research is that performance appraisal cannot be adequately understood outside of its' organizational context, and that the same appraisal system, or the same criteria for evaluating ratings, or the same rater training program, etc. are <u>not</u> the same if they exist in different contexts. "Context" refers to a heterogeneous mix of factors, ranging from the social and legal system in which the organization exists to the climate and culture within the organization. Proximal factors are those that impinge directly on the individual rater, while distal factors affect the rater indirectly (e.g., by determining norms for evaluating performance).

Proximal Variables

Starting from Bandura's (1977) Social Learning theory, Napier and Latham (1986) identified two cognitive process that explain why some raters may consider the appraisal exercise as futile and therefore, perform the task with less rigor. Among these sources is the rater's feeling of self-efficacy, i.e. the degree to which the rater believes he has the necessary skills to perform the task well. Therefore, self-efficacy, as perceived by the individual, would play a motivational role and influence behavioral choices, the mobilization of efforts and the perseverance with which goals are pursued (Fraye & Latham, 1987). Consequently, it is possible to believe that a rater with a high level of self-efficacy would perform the task of appraising the ratee's performance more conscientiously than a rater who does not perceive himself to be able to perform such a task.

The second source of futility is based on outcome expectancies, i.e. the degree to which the rater believes his efforts will be rewarded by the environment. Therefore, a rater who perceives that his task will have few real consequences might very well abandon his desire to carry out performance appraisals of high psychometric quality and his appraisals will be more inclined to be affected by various bias errors (halo effect, central tendency and leniency).

A third proximal variable that deserves scrutiny is the level of confidence in the appraisal system. In addressing this concept, Bernardin and Orban (1985) note that the way in which performance appraisals are carried out may be influenced by a rater's perceptions about the direction in which others bias their performance appraisals. For example, if a rater believes that other raters inflate their ratings to increase the benefits accruing to their subordinates, he might be likely to do the same. Furthermore, the results of research by Bernardin and Orban (1985), performed in a police department, show that when performance appraisals are used for administrative purposes, a low level of confidence in the appraisal system is associated with biased appraisals, i.e., influenced by leniency errors.

Another important class of proximal context variables are those that describe inter-personal relationships between raters and ratees. For example, many researchers have examined the role of affect in performance appraisal. Affect is typically regarded as a potential source of bias in appraisals (e.g., Landy & Farr, 1980; Morin & Dolan, 1992; Tsui & Barry, 1986), but some studies (e.g., Cardy & Dobbins, 1986) suggest that affect toward the ratee significantly influences appraisal accuracy. To date, the role of affect has not been considered in relation to other proximal and distal context variables. Our research will allow us to determine whether other aspects of the rating context might moderate or mediate the effects of affect on appraisals.

Finally, Cleveland and Murphy's (1992) model suggests that raters differ in their beliefs regarding the consequences of giving high or low ratings. For example, some supervisors are likely to be concerned that giving low performance ratings, even where they are clearly deserved, will adversely affect their relationships with the employees who receive such ratings. Raters differ considerably in their beliefs regarding the effects of high or low ratings on the motivation and future performance of their subordinates (Murphy & Cleveland, 1991); raters who view high ratings as a motivational tool may be inclined to give them, even if the employee's performance is in fact poor.

Distal Variables

Murphy and Cleveland (1991) underlined the lack of research on the relationship between organizational climate and performance appraisal quality. Research generally has shown that climate influences the attitudes, behavior and performance of individuals in the organization (Kaczka & Kirk, 1968; Litwin & Stringer, 1968; Pritchard & Karasick, 1973; Waters, Roach & Batlis, 1974). Tziner and Dolan's (1984) research involving real estate agents, shows that the perception of certain dimensions of organizational climate has an effect on performance. More specifically, they found that the more the agent perceived that his environment allowed him to be autonomous, the higher his sales were. This research indicates that there could be a link between the perception of a certain type of climate and performance at work.

Research by Litwin and Stringer (1968) also shows a link between organizational climate and performance at work. Their work is a good starting point to clarify the relationship between organizational climate and performance appraisals of high psychometric quality. These authors conclude that individuals having a high performance need look for organizations with organic climates.

In Tziner and Dolan (1984), as well as Litwin and Stringer (1968), the types of climates described as being linked to high work performance are closely related to Likert's participative model (1961) which is characterized by individual responsibility, cooperative relationships and high performance goals. Likert's work clearly indicates that this type of climate contributes to confidence and loyalty in the work place, good upward, downward and lateral communication, as well as to favorable attitudes between group members. Likert also showed that this type of climate emphasizes feedback and growth of the individual rather than control and punishment.

It is therefore plausible to hypothesize that the closer the climate perceived by an individual is to Likert's participative model and the more a rater perceives that he is in a climate where the appraisal will be used for feedback rather than control, the freer the rater will be to provide appraisals that are accurate and of high psychometric quality. Conversely, the more a rater perceives the organizational climate to be removed from Likert's participative model, the more likely he will be to provide appraisals which are inaccurate and of poor psychometric quality in order to avoid negative consequences for himself and for the ratee.

The second distal variable is to be studied is rater commitment to the organization. Commitment to the organization can be considered in terms of an attitude predisposing to certain types of behavior (Mowday, Steers & Porter, 1979). Commitment can therefore be characterized by the following three factors: (a) a strong belief in and acceptance of the goals and values of the organization, (b) the desire to put forth considerable efforts for the good of the organization, and (c) a strong desire to continue to be a member of the organization. There is a distinction between instrumental commitment and attitudinal commitment (Etzioni, 1961; Gould, 1979; Kelman, 1961; Salancik, 1977; Staw, 1977). Instrumental commitment can be seen as an actor's tendency to pursue in a regular and sustained fashion a series of activities within an organization after having evaluated the costs and benefits (Becker, 1960). As for attitudinal commitment, Meyer and Allen (1984) describe it as the desire, for an employee, to remain within an organization

because this fulfills an intrinsic need related to his personal goals and values. Both forms of commitment might be related to rater behaviors; in a later section of this paper, we describe a proposed model that distinguished between the two types of commitment in relation to several indices of rater behavior.

A few studies have suggested a link between employee commitment and work performance. For example, Mowday et al. (1979) suggest that employees who are highly committed to their organization perform better than employees who are less committed. Based on these empirical results, it would be possible to extrapolate that raters who are more committed to the organization will carry out performance appraisals more conscientiously than raters who are less committed, and consequently, psychometric quality and accuracy should be higher among highly committed raters.

A third distal variable examined in this research involves raters' beliefs about the way in which performance appraisals are used in organizations. There is a substantial body of research showing that when raters believe that appraisals are used to make administrative decisions (e.g. promotions, salary), they are more likely to be lenient than when they believe that ratings are used for feedback, or for some other administrative purpose (for reviews, see Cleveland, Murphy & Williams, 1989; Landy & Farr, 1983; Murphy & Cleveland, 1991). When raters differ in their beliefs about the uses of performance appraisal in their organizations, they may also follow different rating strategies.

Rating Behaviors Affected by Proximal and Distal Influences

The attitudes and beliefs described above might affect a number of aspects of the appraisal process (e.g., how ratings are done, how feedback is handled). Murphy and Cleveland (1995) note that adequate measures of the psychometric quality and accuracy of performance ratings obtained in field settings are extremely difficult to obtain. However, it is possible in most settings to examine: (a) the extent to which raters discriminated among ratees, (b) the extent to which raters discriminated among different aspects of performance, and (c) the extent to which raters assigned high vs. low ratings to their subordinates. These three rating behavior measures can, in turn, be logically and empirically related to the proximal and distal context factors described above.

As noted above, raters who have little trust in the appraisal system, low commitment to the organization, etc. may be less likely to give ratings that clearly discriminate good from poor performers. Different sets of attitudinal variables might lead to different explanations for this effect (e.g., low levels of trust might make raters unwilling to discriminate, whereas low levels of commitment might make them unwilling to invest the effort needed to accurately discriminate), but in general, we expect that several of the distal and proximal beliefs outlined above will affect the rater's willingness to discriminate good from poor performers.

As with discrimination among ratees, different sets of attitudes and beliefs might lead to a lower level of willingness or ability to make these discriminations. As Cleveland and Murphy (1992) note, there are a number of reasons for raters to give all subordinates high ratings,

especially when they have low levels of trust, commitment, self-efficacy, etc.. We expect, for example, that raters' attitudes about performance appraisal will be related to their overall tendency to assign high vs. low ratings to their subordinates.

A Model of The Relationships Between Context Factors and Rating Behavior

The purpose of this study is to examine the network of relationships among several context variables and rating behavior variables described earlier. Table 1 contains predicted relationships among eleven proximal and distal variables that will be measured in the research currently underway. More important, it illustrates the predicted relationships of each of these variables with each of three indices of rater behavior that will be collected in various sites where this research is being conducted. With the exception of instrumental commitment, all eleven variables are predicted to have some direct impact on rater behavior (the effects of instrumental commitment are hypothesized to be indirect.

Table 1 represents the construct explication phase of our research. Pilot studies are currently underway to confirm that the measures used conform to our current understanding of the various constructs involved, and to sharpen our hypotheses about the interrelationships among context and rating behavior measures. Our next step will be to collect data in multiple organizations, spanning national and linguistic boundaries (currently, data collection is planned in the U.S. and in French Canada), to test and further elaborate the model developed and described here.

It is still an empirical question whether the attitudes studied here will prove potent predictors of the rating behaviors being studied. Preliminary results (e.g., Tziner et al., In press) are promising, and we expect that the research that is guided by this model will provide a fresh perspective on why raters give high ratings, fail to discriminate among ratees, etc.. In the very least, the model described here suggests a class of questions that have not, to date, been adequately considered by researchers and practitioners in performance appraisal.

Table 1 - Hypothesized relationships Among Variables in Model

	1	2	3	4	5	6	7	8	9	10	11	12	13
Proximal and Distal Influences													
1. Participative Climate													
2. Instrumental Commitment	#												
3. Attitudinal Commitment	#	+											
4. Self-Efficacy	+	#	#										
5. Confidence in PA	#	#	#	+									
6. Between-People Uses	+	#	+	0	#								
7. Within-People Uses	#	0	+	0	#	0							
8. Organizational Uses	0	+	+	0	+	0	0						
9. Perceived Consequences	-	-	~	-	#	#	+	0					
10. Discomfort with PA	-	-	~	~	~	-	+	0	#				
11. Affect toward Ratee	+	+	+	0	+	0	0	0	0	0			
Rating Outcomes													
12. Rating Level	+	0	-	-	-	+	-	-	-	-	+		
13. Discrimination- Ratees	+	0	+	+	+	-	-	0	-	-	-	-	
14. Discrimination - Dimensions	0	0	+	+	~	+	+	0	+	-	+	0	-

+ weak positive; # strong positive; - weak negative; ~ strong negative

References

Bandura, A. (1977). Social learning theory. Englewood Cliffs, NJ: Prentice Hall.

Becker, H. (1960). Notes on the concept of commitment. American Journal of Sociology, 66, 32-42

Bernardin, H.J. & Orban, J. (1985). Leniency effect as a function of rating format, purpose for appraisal and individual differences. Presented at Annual Meeting of the Academy of Management, Boston.

Cleveland, J.N. & Murphy, K.R. (1992). Analyzing performance appraisal as goal-directed behavior. In G. Ferris and K. Rowland (Eds.), Research in personnel and human resources management, (Vol. 10, pp. 121-185). Greenwich, CT: JAI Press.

Cleveland, J., Murphy, K. & Williams, R. (1989). Multiple uses of performance appraisal: Prevalence and correlates. Journal of Applied Psychology, 74, 130-135. Etzioni, A. (1975). A comparative analysis of complex organizations (rev. ed) New York:Free Press.

Fraye, L.A. & Latham, G.P. (1987). Application of social learning theory to employee self-management. Journal of Applied Psychology, 72, 387-392.

Gould, S. (1979). An equity-exchange model of organizational involvement. Academy of Management Review, 4, 53-62.

Kelman, H. (1961). Process of opinion change. Public Opinion Quarterly, 25, 57-78.

Landy, F.J. & Farr, J.L. (1980). Performance rating. Psychological Bulletin, 87, 72-107.

Landy, F.J. & Farr, J.L. (1983). The measurement of work performance. New York: Academic Press

Likert , R. (1961). New patterns of management. New York: McGraw-Hill.

Litwin, G.H. & Stringer, R.A. (1968). Motivation and organizational climate. Boston: Harvard University Press.

Morin, D. & Dolan, S. (1992). The effect of affect in performance appraisal: A replication and extension of Tsui and Barry's (1984) study. Unpublished manuscript, University of Montreal.

Mowday, R.T, Steers, R.M. & Porter, L.W.(1979). The measurement of organizational commitment. Journal of Vocational Behavior, 14, 224-227.

Murphy, K.R. & Cleveland, J.N. (1991). Performance appraisal: An organizational perspective.Boston: Allyn & Bacon.

Murphy, K. & Cleveland, J. (1995). Understanding performance appraisal: Social, organizational and goal-oriented perspectives. Newbury Park, CA: Sage.

Napier, N. & Latham, G. (1966).Outcome expectancies of people who conduct performance appraisals. Personnel Psychology, 39, 827-837.

Salancik, G.R. (1977). Commitment and the control of organizational behavior and belief. In B. Staw and G. Salancik (Eds.), New directions in organizational behavior. Chicago:St-Clair Press.

Staw, B. (1977). Two sides of commitment. Presented at the Annual Convention of the Academy of Management. Orlando.

Tsui, A.S. & Barry, B. (1986). Interpersonal affect and rating errors. Academy of Management Journal, 29, 586-599.

Tziner, A. & Dolan, S. (1984). The relationship of two sociodemographic variables and perceived climate dimensions to performance. The Canadian Journal of Administrative Sciences 1, 272-287.

Tziner, A., Murphy, K. & Cleveland, J. (In press). Impact of rater beliefs regarding performance appraisal and its organizational contexts on appraisal quality. Journal of Business and Psychology.

Waters, L.K., Roach, D. & Batlis, N. (1974). Organizational climate dimensions and job-related attitudes. Personnel Psychology, 27, 465-476.

Evaluation of a Reengineered Performance Appraisal and Reward System Within the Federal Government [2]

Steven R. Frieman, Ph.D.
Western Area Power Administration

Abstract

This paper describes the evaluation of a reengineered performance appraisal and reward system at a Federal agency of approximately 1400 employees. The design resulting from this effort refocused the performance appraisal system solely on employee development, as well as allowing innovative mechanisms for rewarding individuals, teams, and organizational achievement. The results demonstrate that while current behavioral systems can be adapted to the needs of today's workforce, follow up evaluations are needed to fine-tune them to the organization.

Recent events in the Federal Government, such as Vice President Gore's National Performance Review (1993), have led to unprecedented authority for Federal agencies to make substantial changes in the way their behavioral systems are designed and function. This paper describes the evaluation of a performance appraisal and rewards system which was reengineered and brought on-line in October, 1994. The philosophical and practical considerations involved in this reengineering effort have been discussed in Frieman (1994).

The primary goal of the performance appraisal system reengineering is to have it focus exclusively on employee feedback and development. The redesign accomplishes this goal in three ways. First, it removes status as a consideration by allowing the supervisor to only rate an employee as "pass" or "fail". Second, it removes all rewards from the performance appraisal system. Finally, it adds in a 360-degree feedback process to gather high quality information on how the employee was getting the job done on four critical performance dimensions.

The goals of a redesigned rewards system are to provide meaningful rewards for individuals and teams that exceed performance expectations, to provide these rewards as close in time to the actual performance as possible, and to provide incentives for achieving organizational-wide goals. The redesign effort focused on removing barriers to rewarding groups of employees and allowing supervisors to provide meaningful monetary rewards throughout the year as high performance occurs. The goal of providing incentives to meet organizational-wide goals was met by introducing a Bonus program which only offered a payout if measurable strategic organizational goals were achieved. The intention is that employees will begin to see the links between their job and overall organizational goals, and as a result be more motivated to help in the achievement of such goals.

The ideas presented in this paper are the author's own and do not necessarily reflect the official policy of the Western Area Power Administration.

Method

A survey was conducted of all 1400 Federal employees in this Federal agency during November, 1995 to determine their degree of satisfaction with the Performance Appraisal and Recognition System (PARS) program overall and its individual components. Respondents were asked to rate their overall satisfaction with each component of PARS as well as to indicate what is working well and what needs to be changed. The numerical rating scale ran from 1 through 5, with 1 being "Highly Dissatisfied", 3 being "Neutral", and 5 being "Highly Satisfied." An additional question was added to determine the level of employee support for eliminating second-level review of performance ratings. Feedback on the program was also gathered through a series of employee group interviews.

Results

The survey response rate was 23% Western-wide (324 responses out of 1408 surveys sent out). The response rate ranged from a low of 17% at the CSO to a high of 29% in UGP. Table 1 shows the results of the numerical ratings.

Table 1. Average satisfaction ratings on the PARS system overall and its component parts.

Survey Question	Average Rating
Overall satisfaction with PARS	3.0
Overall satisfaction with the PARS 360 degree feedback system	2.7
Overall satisfaction with the PARS performance appraisal process	3.0
Overall satisfaction with the PARS Special Outstanding Achievement Reward (SOAR) awards process	3.3
Overall satisfaction with the PARS Western Bonus program	3.6

The numerical ratings indicate that employees were "neutral" regarding their overall satisfaction with PARS and its components. Depending on one's viewpoint, responding employees were either slightly satisfied or slightly dissatisfied with the current PARS system. The lowest rating (2.7) was given to the 360 Degree system, and the highest (3.6) to the Western Bonus program. Opinion on eliminating second level review was evenly split at 38% for and against, with remaining responding employees (25%) having no opinion.

Most responding employees were either "neutral" (29%) or "Satisfied" (33%) on the question of whether the PARS program is working well (satisfied) or not (dissatisfied). A remaining 30% were dissatisfied with the program. Considering this was the first year of the program the ratings were acceptable. Other factors possibly influencing the rating of overall

satisfaction with PARS is the amount of new behavior/culture change required such as introduction of a mandatory 360 system, reduction in rating levels, and allowing performance awards throughout the year. In addition, employees feelings may have influenced by an organizational re-engineering process currently in progress.

There seems to be no consensus among responding employees as to whether or not to eliminate second level review. Those in favor of elimination believe it will streamline the rating process and is a non-value added function. Those in favor of retaining it see it as a critical "check and balance" in the performance rating process and helps to keep the second-level aware of how lower level employees are performing.

Many of the responding employees (44%) were in the dissatisfied range with the 360 degree program, while 33% were in the satisfied range. This program received the lowest ratings of any portion of PARS, with an average of only 2.7. It also had the most extreme average ratings, with a low of 2.4 in one area and a high of 3.2 in another. Generally the 360 program is credited with promoting good feedback, open communication, and giving the supervisor a better feel with what the employee is doing. That is, the approach of having 360 feedback is seen as a positive step forward. At the same time there are fundamental questions as to what the final format and process should be. Some of the criticisms of the current process were: 1) that it did not encourage honesty since responses were not anonymous; 2) there was too much paperwork in the system; 3) the process takes too much time (to collect feedback); 4) the feedback from supervisors was not constructive; and 5) the concept of 360 may be redundant for line crews who continually work in a team environment.

Proposed suggestions for change from responding employees were: 1) make the 360 process anonymous to encourage more honest responses; 2) eliminate excessive paperwork; 3) exempt line crews from the 360 process since they receive immediate feedback on job performance working as a natural team; and 4) provide more training to supervisors and managers on how the system works and how to use it effectively to provide constructive feedback.

Most responding employees were neutral (30%) or satisfied (31%) with the performance appraisal portion of PARS. Generally the process was seen as more streamlined and less paperwork intensive than the previous, traditionally designed, system. Another plus was the delinking of awards from the rating process. The major change suggested was to move to a pass/fail rating system by eliminating the Outstanding level rating. This particular change has already been put into place.

45% of responding employees were either highly satisfied or satisfied with the Special Outstanding Achievement Reward (SOAR) process, another 25% were neutral, and approximately 22% were in the highly dissatisfied or dissatisfied range. Generally the awards process is seen as simplified and easy to use. Real positives are the ability of others to initiate nominations, the increased frequency of awards and the closeness in time of the award to the actions being rewarded. The major criticism of the SOAR program revolved around supervisor understanding of how it operated, especially in terms of the criteria to approve an award. There needs to be clarification as to when a SOAR award is appropriate. For example, should it be

given for simply "doing one's job" or exclusively for an exceptional event, should it go to the "star" performer or to the support staff, should it go to supervisors or only non-supervisory employees, etc. This confusion has led to some supervisors waiting for others to initiate a nomination instead of doing it themselves.

Statistical information on the SOAR program supports the above survey information. Since SOAR was introduced spending for all awards has decreased to 1.3% of total salary from 1.5% in prior years, while the number of awards has increased. That is, employees were satisfied with the awards process, even with smaller award amounts, because the awards were more frequent and targeted to specific achievements. Had the SOAR system been in place in prior years it would have resulted in an awards spending savings of $150,000 for each of FY 93 and FY 94. (Note: FY 95 Bonus payout costs are not included in the above comparison). In addition, 43% of all approved SOAR nominations were initiated by someone other than the immediate supervisor of the program. This means that almost half of all SOAR awards was for employee actions that the supervisor did not see, but that the supervisor agreed was worthy of an award.

The Bonus program was the only portion of PARS with a majority of responding employees (55%) in the satisfied or above range. Only 12% of employees were dissatisfied in any way with the program. The Bonus program is seen as a positive step forward in creating a "One Organizational" culture. It was seen as promoting teamwork and being objective enough for all employees to understand. There were three major criticisms of the program. First, employees are not connected to the program and do not understand how they can impact it. Suggestions to rectify this included involving employees in the goal setting process, providing more information on how employees can impact goals, and having some regionally based goals. Secondly, there is little or no reinforcement of the program throughout the year. Finally, the eligibility requirements are unfair in that part-time employees are prorated for their time worked, while new full-time employees are not, even if they have been on board less than a year.

Discussion

Overall the PARS program is seen as a positive step forward and should be retained. However, specific components of the PARS system will require some fine-tuning as detailed below.

The performance appraisal process is working well at this time and no recommendations for changes in the rating process are proposed. The goal of refocusing it on employee development has been achieved, although there is dissatisfaction with the 360 Degree feedback component of the program. It is unclear from the survey whether the dissatisfaction is the format, process, cultural change or some combination of these factors. Further investigation of the 360 process is needed to identify the root causes of the dissatisfaction and make recommendations for specific changes.

The overall SOAR program is working exceptionally well in its first year. No structural changes are recommended. It even appears that the SOAR program is helping to contain costs with its emphasis on more frequent awards in smaller amounts. The SOAR award criteria was

written broadly to allow for local fine-tuning as needed. This apparently caused some confusion as to the conditions under which a SOAR award was appropriate. Local units will need to clarify within management ranks the conditions under which a SOAR award is to be approved, especially in the areas of: 1) what actions may make a supervisor eligible for an award; 2) what actions may make a low profile employee eligible for an award (e.g., with a support staff or assistant position); 3) what actions may make a team or group of employees eligible for an award.

The future success of the Bonus program requires that employees are connected to the goals of the program. Connection comes from involvement and priority. With this in mind the following recommendations are proposed. First, solicit employee suggestions for Bonus program goals for the FY 97 Bonus year by April, 1995. Second, have management at all levels of the organization discussing the program on a regular basis with employees in their organization. Finally, improve written communication throughout the year with simpler monthly progress reports and more focus on what employees can do, or are doing, to help achieve the goals.

In conclusion, the reengineered performance appraisal and recognition systems performed well in their first year of operation. At the same time it is critical for yearly evaluations of these behavioral systems to take place in order to more accurately adjust them to meet the needs of a Federal employee workforce. Through the use of periodic evaluations, a continuous process improvement strategy can become integrated with policy formulation and implementation.

References

National Performance Review (1993). From red tape to results: Creating a government that works better & costs less (GPO No. 040-000-00592-7). Washington, D.C.: U.S. Government Printing Office.

Frieman, S. (1994). Revising performance appraisal and reward systems within the Federal government. Proceedings of the 14th Applied Behavioral Science Symposium, 156-159.

The Relation of Prior Performance Feedback Ratings to Managers' Subsequent Feedback Seeking Behavior

Ann M. Herd, Ph.D.
United States Air Force Academy

Abstract

The feedback seeking behavior of 153 midlevel managers participating in a developmental assessment workshop was investigated in relation to feedback ratings given by each manager's supervisor, subordinates, and self prior to the workshop. Results indicated, as hypothesized, that managers sought feedback more when prior ratings were lower. In addition, they used a supervisor monitoring strategy more when their supervisor's rating was lower than their self ratings. However, this finding did not hold true for the subordinate monitoring strategy. In addition, the prediction that managers would use an inquiry strategy more when others' ratings were higher than their self ratings was not supported.

Historically, feedback has primarily been studied as an organizational resource, due largely to the performance-enhancing effects of feedback. In recent years, however, researchers have suggested that feedback also serves as an individual resource which employees are actively motivated to seek (Ashford & Cummings, 1983). While a number of recent studies have investigated factors related to employees' feedback seeking behavior, areas which have received relatively little attention include the feedback seeking behavior of managers as a group, and the effects of prior feedback on the feedback seeking process.

The feedback seeking behavior of managers as a group may be a particularly important area for study because managers' needs for feedback from others may be greater than that of other employee groups. The performance dimensions for managerial jobs are generally less concrete than those for other job types, and managers' job tasks often do not inherently provide feedback. In addition, much of a manager's success involves the ability to deal with supervisors, coworkers, subordinates, and clients. Thus, managers may be particularly cognizant of the need for feedback about their performance from others in their work environment.

One study which did investigate managers' feedback seeking behavior found that managers who sought negative feedback from various sources (e.g., supervisor, peers, subordinates) had a more accurate understanding of others' perceptions of their work than managers who did not seek negative feedback (Ashford & Tsui, 1991). In addition, these sources of feedback had more positive views of managers who sought negative feedback and more negative views of managers who sought positive feedback.

The present study was designed to investigate managers' feedback seeking behavior in relation to prior feedback ratings from their supervisor and subordinates, as well as prior ratings they gave themselves. While no study has investigated these variables specifically, research on the motives for

seeking feedback as well as on feedback seeking strategies and their associated costs can provide direction for hypothesized effects.

Researchers suggest three categories of feedback seeking motives (Ashford & Cummings, 1983; Levy et al., 1995). One motive for seeking feedback is goal attainment; that is, feedback provides information for individuals to help meet their goals, and reduces uncertainty regarding whether and how these goals can be achieved. A second motive affecting individuals' feedback seeking behavior is the desire to protect one's ego. This motive would suggest that individuals may be more inclined to seek positive rather than negative feedback, or to seek negative feedback only in such a way that the ego is not challenged. A third motive affecting feedback seeking behavior is that of impression management, or the desire to present oneself favorably to others.

These motives are closely tied to the choice of feedback seeking strategy and the costs perceived as associated with this choice (Ashford & Cummings, 1983). One strategy, inquiry, consists of directly asking a source for feedback. Inquiry is usually perceived as entailing potential face-loss and inference costs for the seeker. On the other hand, a monitoring strategy involves paying attention to and interpreting information available in the feedback environment (e.g., watching the source's facial signals and reactions to one's performance). A monitoring strategy is usually perceived as potentially entailing greater effort costs than inquiry.

In predicting managers' feedback seeking responses to prior feedback, the ego defensive motive would suggest that managers may subsequently avoid seeking feedback from a source whose prior feedback was negative. On the other hand, goal attainment and impression management opportunity motives may suggest that managers would seek more feedback after receiving negative feedback from a source. This feedback could help to reduce uncertainty and obtain valuable information about how to improve their performance. This feedback seeking could also provide the opportunity to influence the source's impressions in a positive manner. Since the managers in the present study were working to achieve personal developmental goals which they themselves chose, it was hypothesized that goal attainment motives would be stronger than ego-defensive motives in predicting managers' feedback seeking behavior.

Hypothesis 1: Prior feedback ratings from a particular source will be negatively related to subsequent feedback seeking from that source.

In predicting managers' choice of feedback seeking strategy, it was hypothesized that the managers' self-assessments may provide an "anchor" or standard by which managers may judge the relative sign of feedback ratings from another source. Ratings which are more discrepant from the manager's self rating may cause increased uncertainty, which may result in increased overall motivation to seek feedback. The sign of this discrepancy would seem an important factor in strategy choice. For example, negative discrepancies (where others' ratings are lower than self ratings) may elicit greater perceptions of face loss costs, while positive discrepancies (where others' ratings are higher than self ratings) may engender less face loss costs.

Hypothesis 2: Individuals who have received negatively discrepant feedback from a given source will use a monitoring strategy with that source more than will individuals who have received positively discrepant feedback from that source.

Hypothesis 3: Individuals who have received positively discrepant feedback from a given source will use an inquiry strategy with that source more than will individuals who have received negatively discrepant feedback from that source.

Method

Sample

The sample consisted of 153 middle-level managers in a large governmental agency who participated in a developmental assessment center (M age = 47 years, SD = 7.6; M organizational tenure = 18 years, SD = 8.2; M job tenure = 5 years, SD = 3.6). In all, 266 managers who had participated in the workshop were sent a follow-up questionnaire containing the measures in the present study. A total of 157 questionnaires were returned, yielding a response rate of 59%. Of these, 153 questionnaires were useable.

Procedure

Subjects in the study participated in an agency-wide, four-day developmental assessment center (called the Skills Assessment Workshop, or SAW) required for all managers at all levels of the agency. The formal objectives of the workshop were for participants to: receive feedback on specific managerial competencies from supervisor's, peers', subordinates', and self assessments; identify strengths and areas for improvement; and write an Individual Development Plan consisting of specific goals and actions.

Approximately one month prior to the workshop, participants were sent a packet of questionnaires to be distributed as follows: one questionnaire was to be given to his/her supervisor, one questionnaire was to be completed by the participant him/herself, and the other five questionnaires were to be distributed to his/her subordinates for their feedback ratings. Each questionnaire contained 60 items assessing the proficiency level of the manager on 16 managerial skill dimensions (e.g., oral communication, written communication, problem solving and analysis, developing subordinates, etc.). Questionnaire instructions directed the raters to complete the questionnaires and return them to the training department within two weeks. Supervisors were assured their ratings would be used only for developmental purposes, and would be kept private between themselves and their rated subordinate. Subordinates were likewise informed in the questionnaire instructions that their ratings would be averaged with other subordinates' ratings, and so would be confidential.

Once all the feedback questionnaires had been sent back to the training department, the results were tabulated for each of the workshop participants. These results were given to the participants as feedback on the third day of the workshop. Feedback from the three sources was presented in the form of averaged dimension ratings from all sources as well as averaged individual item ratings from all sources. At the end of the workshop, participants completed an Individual Development Plan (IDP), where they identified their three highest priority developmental objectives, as well as an action plan regarding how they planned to meet their developmental objectives.

Approximately three months after the workshop, participants were sent a follow-up questionnaire on which they were asked to identify their most important IDP goal and answer questions regarding their progress toward this goal (including most of the measures for the present study).

Measures

Proficiency Ratings. For purposes of the present study, the proficiency feedback ratings used for analysis included only those pertaining to the participant's reported most important IDP goal (measured by the pre-workshop questionnaire described above). For each performance dimension, raters were asked to rate the focal manager on several items using a five-point proficiency scale (1 = very low level to 5 = very high level). For each subject, the mean rating for the items constituting the most important goal was used as the proficiency rating measure in the present study.

Feedback Seeking Behavior and Reliance. Various aspects of managers' feedback seeking behavior and reliance were the dependent variables of interest in the present study. Altogether, 17 items were combined in various ways to measure the following feedback seeking and reliance variables: overall feedback seeking behavior, supervisor inquiry, supervisor monitoring, subordinate inquiry, and subordinate monitoring. Ten of the items were the same as those used by Ashford (1983). Seven items were adapted from Ashford's items to assess subordinates as a source of feedback seeking. For each feedback seeking item, subjects were asked to rate the frequency with which they engaged in the feedback seeking behavior, on a 5-point Likert scale ranging from 1 = "Very Infrequently" to 5 = "Very Frequently". The mean rating on these items yielded a scale score ranging from 1 to 5, with higher values indicating more frequent feedback seeking behavior.

Results

Descriptive statistics for variables in the study are shown in Table 1, while correlations for variables in the study are presented in Table 2.

As shown in Table 2, the correlation between supervisor ratings and overall feedback seeking from the supervisor was significant ($r=.21$, $p<.05$). Likewise, the correlation between subordinates' rating and overall feedback seeking of subordinates was significant ($r=.18^*$, $p<.05$). As predicted, the direction of these correlations was negative. Further exploration of seeking strategy revealed negative correlations between source ratings and use of a monitoring strategy with that source, but less strong (although still negative) correlations between ratings and inquiry.

To test the second and third hypotheses, subjects were divided into two groups: those receiving negatively discrepant feedback from a source (i.e. their self rating was higher than the source's rating of them), and those receiving positively discrepant feedback from a source (i.e. their rating was lower than the source's rating of them). A t-test comparing the mean supervisor monitoring values of the two groups revealed significant differences in the monitoring of the two groups for the supervisor as a feedback source ($t=2.91$, $R^2=.06$, $p<.01$). Examination of the mean supervisor monitoring scores of the two groups revealed that subjects in the negatively discrepant group ($N=72$) monitored their supervisor for feedback more ($M=3.21$, $SD=.75$) than did subjects in the positively discrepant group ($N=56$,

Table 1. Descriptive Statistics for Variables in the Study.

Variable	M	SD	Actual Range	α
Prior Feedback Ratings				
Supervisor's Rating	3.50	.67	1.0 - 5.0	
Subordinates' Rating	3.49	.58	2.1 - 5.0	
Self Rating	3.60	.52	2.3 - 5.0	
Feedback Seeking (FSB)				
Overall Supervisor FSB	2.68	.76	1.0 - 4.6	.85
Supervisor Inquiry	2.11	.87	1.0 - 4.5	.78
Supervisor Monitoring	2.97	.86	1.0 - 5.0	.80
Overall Subordinate FSB	2.89	.84	1.0 - 4.4	.89
Subordinate FSB Inquiry	2.49	.98	1.0 - 4.5	.79
Subordinate FSB Monitoring	3.05	.88	1.0 - 4.8	.84

Table 2. Zero-Order Pearson Correlations Among Measures in the Study.

Variable	1	2	3	4	5	6	7	8	9	*
1. Supervisor's Rating	40***	18*	-21*	-13	-25*	-22*	-16	-25**		
2. Subordinates' Rating	--	22*	-02	-03	-06	-18*	-08	-21**		
3. Self Rating		--	-03	-04	-02	-09	-04	-08		
4. Overall Supervisor FSB			--	81***	92***	59***	48***	59***		
5. Supervisor Inquiry				--	56***	49***	58***	41***		
6. Supervisor Monitoring					--	56***	37***	63***		
7. Overall Subordinate FSB						--	86***	94***		
8. Subordinate FSB Inquiry							--	65***		
9. Subordinate FSB Monitoring									--	

Note: Decimal points have been omitted.
* $p < .05$, ** $p < .01$, *** $p < .001$.

M=2.79, SD=.85). For subordinates as a source of feedback, a t-test comparing the mean subordinate monitoring scores of subjects receiving negatively discrepant feedback from their subordinates (N=81) versus those receiving positively discrepant feedback from their subordinates (N=50) revealed no significant difference in subordinate monitoring between the two groups. Thus Hypothesis 2 was supported for supervisors as a source of feedback but not for subordinates.

For the inquiry strategy, a t-test comparing the mean supervisor and subordinate inquiry scores of subjects in the negatively discrepant groups versus subjects in the positively discrepant groups revealed no significant difference between the two groups. Thus, Hypothesis 3 was not supported.

Discussion

Results of the study suggest that managers' feedback seeking behaviors (particularly using the monitoring strategy) are increased when feedback ratings from others are lower. Given that negative feedback ratings may pose more threats to the ego, these findings suggest that ego-defensive motives may have been less salient to the managers in this study than goal-achievement or impression management motives for seeking feedback. Study results overall did not support hypotheses regarding the sign of discrepancy between managers' self ratings and others' ratings as a predictor of subsequent feedback seeking behavior.

Since the performance feedback managers obtain ultimately may affect their performance and achievement of goals, future research like the present study is needed to facilitate increased understanding of managers' feedback seeking behavior from various sources in their environment, and their choice among various motives, sources and strategies for various types of feedback.

References

Ashford, S.J. (1983). Coping with uncertainty: Feedback seeking in a changing environment. Unpublished dissertation.

Ashford, S.J. & Cummings, L.L.(1983). Feedback as an individual resource: Personal strategies of creating information. Organizational Behavior and Human Performance, 32, 370-398.

Ashford, S.J. & Tsui, A.S. (1991). Self-regulation for managerial effectiveness: The role of active feedback seeking. Academy of Management Journal, 34(2), 251-280.

Levy, P.E., Albright, M.C., Cawley, B.D., & Williams, J.R. (1995). Situational and individual determinants of feedback seeking: A closer look at the process. Organizational Behavior and Human Decision Processes, 62, (1), 23-37.

Factors Contributing to the Morale, Cohesion, and Motivation of Combat Support Personnel During Desert Shield/Desert Storm

Captain Gary Jandzinski
Dr. David Vaughan
Lt Col Jim Van Scotter

Abstract

Morale, cohesion, and motivation are viewed as prerequisites to the success of military campaigns, but there has been little empirical research investigating their inter-relationships or relationships with other variables. This study examined the influence of situational factors on the cohesion, morale, and motivation experienced by (N=71) USAF aircraft maintenance personnel who participated in Operations Desert Storm/Desert Shield. Results suggest that variables reflecting the quality of the living conditions and social support they received explained significant variance in their morale, cohesion, and motivation.

Morale, cohesion, and motivation are often mentioned as determinants of wartime performance. Military planners and leaders view them as important in sustaining high levels of individual performance under difficult conditions (Borman, Johnson, Motowidlo, and Dunnette, 1979). Accounts of the Gulf War suggest that morale, cohesion, and motivation were each important factors in the Coalition victory (Winnefield, et al., 1994). Unfortunately, most of what we know about them is based on anecdotes, so it is impossible to provide commanders and supervisors clear guidance on how they can increase morale, cohesion and motivation in their units. Our paper reports empirical work that begins to address this problem.

Following Kellet (1986), we define morale as an individual's mental and emotional attitudes towards the duties he or she is expected to perform (Kellet, 1982). It is a sense of individual psychological well-being based on a sense of common purpose and the expectation of successful group performance. Morale is a key factor in groups with high achievement levels (Gal, 1987). The desire for group achievement and successful performance are important factors in morale. In comparison, the central themes of cohesion are group membership, loyalty, shared values, and identification with the group (Shalit, 1988). Cohesion is more group-oriented than-achievement oriented. Unit cohesion plays a large part bolstering individual self-confidence in a combat situation (Gal, 1986). Motivation is described in terms of the direction, intensity, and duration with which an individual pursues his or her goals. Each of these factors is influenced by the other attitudes as well as external quality of life factors.

Our purpose in this study is to investigate the inter-relationships between morale, cohesion, and motivation and to test their relationships with two kinds of situational factors. The first situational factor (living conditions) focuses on the quality of the food, billeting arrangements and related facilities available during the conflict. The second situational factor (social support) centered on the availability and quality of MWR support, entertainment, mail/phone

communication with relatives at home, and information about events in the theater of operations. We hypothesized that living conditions and social support would each have independent effects on the level of morale, cohesion, and motivation experienced by maintenance personnel participating in Desert Shield/Desert Storm.

Methods

Subjects

Aircraft maintenance personnel (N=74) who had been stationed within the Desert Storm/Desert Shield theater of operation between August 1990 and July 1991 participated by completing a survey that measured morale, cohesion, motivation, and two situational factors. Most participants were enlisted (N=70) males (N=71). All of them were stationed in the theater of operations between August 1990 and July 1991.

Instrument

A survey was developed for this study. Three 3-item scales adapted from Gal's (1986) study of Israeli soldiers were used to measure the level of morale, cohesion, and motivation experienced by the subject. Items describing conditions during the war were generated by veterans of the conflict in unstructured interviews. Eight items measured subjects' satisfaction with the living conditions and food quality they experienced during the war. Eighteen items asked subjects about the quality of the MWR support, mail/phone service, entertainment, and information they received about the war through unofficial news sources. Responses to all items used the same 5-point scale. Anchors for the scale ranged from 1=poor to 5= excellent.

Results

The Cronbach's alphas in Table 1 show that measures have adequate internal consistency. The table also provides evidence that morale, cohesion, and motivation are highly related, echoing Gal's (1986) findings.

Morale, cohesion, and motivation were each used as dependent variables in separate hierarchical set regression analyses (Cohen & Cohen, 1983). We followed the same procedure for each dependent attitude variable. The first set of variables entered into the regression was comprised of the two attitude measures that were not serving as the dependent variable. In the next step the set of living condition variables was added to the regression. Then the social support variables were entered as a group. Next the attitudinal variables, and then the living condition variables were removed from the analysis. This procedure calculates the change in the variance explained that is uniquely attributable to the attitude variables, living condition measures, and social support variables. Table 2 shows the results of all three sets of analyses. The change in R2 is shown in the column labeled "change" in Table 2.

Table 1
Intercorrelations Among the Study's Variables

Variable	1	2	3	4	5	6	7	8	9
Attitudes									
1. Morale	(.93)								
2. Cohesion	.58	(.92)							
3. Motivation	.61	.58	(.91)						
Living Conditions									
4. Billeting	.48	.27	.21	(.86)					
5. Food	.45	.30	.19	.58	(.87)				
Social Support									
6. MWR	.47	.36	.34	.39	.47	(.91)			
7. Entertainment	.43	.19	.10	.37	.45	.55	(.74)		
8. Mail	.32	.27	.31	.57	.54	.40	.42	(.76)	
9. Information	.60	.48	.36	.67	.61	.40	.44	.48	(.84)

Notes: N=74. $p<.05$ for $r>.22$, $p<.01$ for $r>.29$ (two-tailed).
Cronbach's alphas shown on the diagonal.

The analyses showed that attitude variables consistently accounted for more variance in the dependent attitude variables than the situational factors did. Living conditions and social support contributed 22 percent of the variance in morale over that accounted for by cohesion and motivation. It is also worth mentioning that nearly 50 percent of the variance in morale is accounted for (not uniquely) by the situational factors in Sets 2 and 3 combined.

Only the set of variables comprised of morale and cohesion accounted for significant incremental variance in motivation, (about 21 percent). However, 26 percent of the variance in motivation is accounted for by Set 2 (food, living conditions) and Set 3 (information, MWR, mail, entertainment) together when Set 1 is removed. As with morale, this large percentage indicates that these variables are important as a composite group of situational factors.

Table 2
Hierarchical Set Regression Results

Dependent Variable: Step Procedure	Morale R2 Change	Cohesion R2 Change	Motivation R2 Change
1 Enter other two attitude variables	.38 ------	.35 ------	.40 ------
2 Add living condition set	.49 (.11*)	.37 (.02)	.41 (.01)
3 Add social support set	.60 (.11*)	.39 (.02)	.47 (.06)
4 Remove attitude predictor set	.50 (-.10*)	.26 (-.13*)	.26 (-.21*)
5 Remove living condition set	.48 (-.02)	.24 (-.02)	.23 (-.03)

Notes: N=74 for all analyses. *p<.10.

Discussion

This study supports the view that morale, cohesion, and motivation are related, but different constructs. The pattern of correlations among the variables provides considerable evidence that morale is more strongly related to external factors than are the other two attitudes. Regression results also showed that living conditions and social support factors have a stronger influence on morale than attitudinal factors do. Morale and cohesion explain significant variance in motivation and morale and motivation explain a significant portion of the variance in cohesion. Results also provide evidence that situational factors make a more important contribution to morale, than to cohesion or motivation. These findings are consistent with the way morale, cohesion, and motivation were defined earlier.

These results hint at a model in which the quality of the living conditions and social support deployed personnel recieve influences their morale, which in turn increases their motivation and cohesion. These relationships suggest that commanders and supervisors may be able to make significant improvements in all cohesion and motivation by improving living conditions and social support in ways that affect morale.

References

Borman, W. C., Johnson, P. D., Motowildo, S. J. and Dunnette, M. D. (1975). Measuring motivation, morale, and job satisfaction in army careers. Alexandria, VA: U.S. Army research institute for behavioral sciences.

Belenky, G. et al. (1987). Contemporary studies in combat psychiatry. Westport, CT: Greenwood press.

Cohen, J., and Cohen, P. (1983). Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, NJ: Marcel dekker, inc.

Gal, R. (1986). Unit morale: from a theoretical puzzle to an empirical illustration an Israeli example. Journal of Applied Social Psychology, 16, 549-564.

Gal, R., and Manning, F. J. (1987). Morale and its components: a cross-national comparison. Journal of Applied Social Psychology, 17, 369-391.

Kellett, A. (1982). Combat motivation: the behavior of soldiers in battle. Boston, MA: Kluwhernijhoff publishing.

Shalit, B. (1988). The psychology of conflict and combat. New York: Praeger.

Winnefeld, J. A., Niblack, P., and Johnson, D. (1994). A league of airmen--U. S. air power in the gulf war. Santa Monica, CA: Rand corporation.

Evidence of the Usefulness of the Trait of Agreeableness for Selecting Employees to Reduce Performance Variability in Critical Group Tasks

Capt Max R. Massey
Lt Col James R. Van Scotter
Guy S. Shane
Air Force Institute of Technology

Abstract

This study tested the hypothesis that the personality trait of agreeableness influences the variability of two-person team performance in a complex, unfamiliar task. After being pretested on agreeableness with Wiggins (1988) Revised Interpersonal Adjective Scale, (N=55) subjects were assigned to two-person teams comprised of members with high scores on agreeableness (N=11), or two-person teams whose members had low scores on agreeableness (N=11). Eleven other subjects participated in the experiment as individuals. Subjects in all three conditions completed an aircraft load planning exercise on five consecutive days. Analysis showed significant differences in the amount of variation in performance across the three conditions. The high-agreeableness teams' performance was less variable than the either of the other groups over the five day trial, although mean performance levels were not significantly different. Total job experience and experience in the work center also explained substantial variance in performance. Results suggest that the trait of agreeableness may be useful in selecting people to work on teams responsible for performing critical or sensitive tasks.

Erratic performance by surgeons, astronauts, pilots, and military personnel in wartime, can lead to loss of life or critical resources. Performance variability increases uncertainty about the outcomes of critical tasks and puts key organizational resources at risk. One way organizations attempt to reduce performance variability is by ensuring that their most critical tasks are assigned to experienced workers. Another way is by assigning critical tasks to teams or small groups. Small teams handle some of the most critical tasks in the military (e.g., missile launch crew). Members of teams working in a hazardous or dangerous situations such as the deck of an aircraft carrier, or explosive ordinance disposal are especially dependent on each other to perform reliably and predictably.

At the individual level of performance, the personality trait of agreeableness is associated with discipline and leadership (Hough, Eaton, Dunnette, Kamp, & McCloy, 1990), helpful, considerate, and compliant behavior (Motowidlo & Van Scotter, 1994) and overall job performance ratings (Tett, Jackson, & Rothstein, 1991). Research also makes it clear that job experience affects the variability of individual performance (Hunter, Schmidt, & Judiesch, 1990; McDaniel, Hunter, & Schmidt, 1988); especially when experience levels are low or tasks are complex (Avolio, Waldman, & McDaniel,. 1990; Schmidt, Hunter, Outerbridge, & Goff, 1988).

Although the amount of variance in performance can mean the difference between success and failure in small work groups, there has been little research to identify factors that influence the amount of variability in small group performance. This study takes a first step by investigating the usefulness of the personality trait of agreeableness in predicting the variability of two-person small group performance on a complex, unfamiliar task. Our primary hypothesis is that the performance of two-person teams whose members have high scores on agreeableness will exhibit less variation than teams whose members had low scores on agreeableness. However, we do not expect mean performance to differ between the groups. A secondary hypothesis is that experience will account for a significant portion of the variance in job performance.

Method

Sample

Subjects for the study were (n=82) enlisted Air Force members assigned to an aerial port. All subjects completed a questionnaire designed to collect demographic data and measure the trait of agreeableness in the first phase of the study; 55 completed all phases of the research.

Instrument

A preexperiment questionnaire was administered about two weeks before the main study began. Wiggins et al's. (1988) Interpersonal Adjectives Scales-Revised (IASR-B5) personality inventory measured agreeableness. Cronbach's alpha for the scale was .82 (N=82) in this sample.

Procedure

In order to obtain estimates of performance variability, we designed a repeated-measures experiment. After we ensured that the subjects had no prior experience with the experimental task, subjects were assigned to one of three groups. Group 1 consisted of eleven two-person teams which were formed from the twenty-two people scoring highest on the dimension of agreeableness (M=77.32,SD=6.21). Group 2 consisted of eleven two-person teams formed from the twenty-two people scoring lowest on agreeableness (M=59.32, SD=4.48). Group 3 consisted of eleven subjects, who participated in the experiment as individuals, completing the same tasks the two-person teams did.

Participants completed five aircraft load-planning tasks, one each day for five consecutive days. Obtaining repeated performance measures over this period made it possible to assess variability in the teams' performance over time. The basic task was the same thoughout the experiment, but the details changed from day to day. Each day the participants were given a consolidated, randomly ordered list of cargo pallets. It contained information on pallet location and other information needed for load-planning including weight, hazardous material class, priority, and how long it had been waiting to be shipped. Participants used this information to prepare load plans.

Group 1 and Group 2 were instructed to complete the scenarios as a team and agree on all responses before recording them. Group 3 was instructed to complete the scenarios individually. All participants were instructed to work without any outside help. The participants were given verbal instructions to record completion time, destination, pallet identification, pallet location, pallet hazard classification, pallet weight, and total cargo weight on answer sheets that were provided. This procedure was expected to encourage them to interact with each other frequently.

The primary response variables were completion time, cargo weight, safety errors, administrative errors, cargo priority, and cargo age. These variables were scored by the researchers after all exercises were complete. Detailed procedures were established to ensure uniformity. Subjects were provided with information about task requirements and objectives, but did not receive any feedback about their performance during the experimental sessions.

Results

The Box-M test was used to test the homogeneity of the variability exhibited by the three groups. The $p<.10$ significance level was used for these analyses because there were less than 20 subjects per group (Stevens, 1992:175). Results for the Box test (Table 1) show that the groups' variances were not homogenous for four of the five criteria. The sixth criterion measure, safety errors, was excluded from the statistical analysis because the base rate was too low to support the analysis. The results support the hypothesis that the performance of groups in which both members have high agreeableness scores would be less variable than the performance of groups in which both members have low agreeableness scores.

Table 1
Groups 1-3 Performance Variability Over Five Trials

	Group 1 High-Agr Variance	Group 2 Low-Agr Variance	Group 3 Individuals Variance	Significance of Criterion Box-M Test
Admin Errors	39.88	56.30	34.58	$p<.03$
Completion Time	196.55	284.21	184.09	$p<.06$
Cargo Age	66.0K	83.0K	79.0K	$p<.08$
Cargo Priority	124.88	135.08	152.28	$p<.01$
Cargo Weight	460.0M	597.0M	394.0M	NS

Notes: N=5 trials for variances. K=thousands;M=millions.
NS=non-significant.

Multivariate Analysis of Covariance (MANCOVA) procedures were used to test the differences in the three groups' mean scores on the criterion measures. Between-group effects were non-significant. Thus, the hypothesis that the high-agreeableness groups, low agreeableness groups, and individual participants did not differ in their aggregate performance over the five trials

could not be rejected. To ensure that including the individual participants and the two person teams did not confound the analysis, the MANCOVA procedure was repeated for just the high- and low-agreeableness groups. The results did not change.

The MANCOVA results also provided information about the relationship of experience to performance on the experimental task. The squared partial correlations shown in Table 2 were obtained in separate MANCOVA analyses in which either work center experience, or total Air Force experience was entered as a covariate. The results suggest that work center experience (which may be a crude index of the subject's previous contacts with each other) explains substantial variance in performance on the three production-oriented measures, whereas total Air Force experience explains substantial variance in compliance-oriented measures (completion time, administrative errors, and safety errors).

Table 2

Squared Partial Correlation for Experience in MANCOVA

Type of Experience	Cargo Weight	Cargo Age	Cargo Priority	Completion Time	Admin Errors	Safety Errors
Work center	.99	.99	.85	.02	.37	.30
Total AF	.01	.01	.15	.98	.63	.70

Notes: N=33 groups over N=5 occasions for all criterion variables.

Discussion

Group performance, like individual performance, varies significantly between occasions. Our analysis supports the hypothesis that agreeableness is associated with differences in variance in two-person team performance, but is not associated with differences in mean performance levels. Four of five Box-M tests identified significant differences in heteroscedasticity among the groups (Table 1). Evidence also supported the influence of experience on performance variability (Table 2). We found that general experience (measured here as total Air Force time) is significantly correlated with variance in compliance-oriented outcomes (completion time, safety errors, and administrative errors), whereas work center experience explains variance in production-oriented outcomes (cargo weight, age, and priority). Thus, the utility of experience for assigning individuals to critical tasks depends on the nature of the task and the nature of the experience.

The results suggest that even when individuals work together on a difficult, relevant task, personality differences and task characteristics and demands all influence behavior. Understanding how this occurs in high-risk situations or situations clearly linked to organizational goals seems especially important for the military. Future research should investigate the influence

of varying degrees of task complexity. Our results support Yetton and Johnston's (1992) argument for the importance of incorporating performance variability in performance theory.

Subjects participating in the present study knew they were being evaluated on a task that was at least indirectly related to their duties in the Air Force. The knowledge that they were being evaluated, even if only for research purposes, may have increased the pressure to perform. If increased pressure to perform was perceived, the affect of agreeableness on group performance should have been attenuated. In this case, the results may be somewhat understated.

References

Avolio, B.J., Waldman, D.A., & McDaniel, M.A. (1990). Age and work performance in non-managerial jobs: The effects of experience and occupational type. Academy of Management Journal, 33, 407-422,

Hough, L.M., Eaton, N.K., Dunnette, M.D., Kamp, J.D., & McCloy, R.A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. Journal of Applied Psychology, 75, 581-595.

Motowidlo, S.J., & Van Scotter, J.R. (1994). Evidence that task performance should be distinguished from contextual performance. Journal of Applied Psychology, 79, 475-480.

Tett, R.P., Jackson, N.D., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analysis. Personnel Psychology, 44, 703-742.

Hunter, J. E., Schmidt, F.L., & Judiesch, M.K. (1990). Individual differences in output variability as a function of job complexity. Journal of Applied Psychology, 75, 28-40.

McDaniel, M.A., Hunter, J.E., & Schmidt, F.E. (1988). Job experience correlates of job performance. Journal of Applied Psychology, 73, 327-330.

Schmidt, F.L., Hunter, J.E., and Outerbridge, A.E. (1986).Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. Journal of Applied Psychology, 71, 423-429.

Schmidt, F.L., Hunter, J.E., Outerbridge, A.E., & Goff, S. (1988). Joint relation of experience and ability with job performance: A test of three hypotheses. Journal of Applied Psychology, 73, 46-57.

Stevens, J. (1992). Applied Multivariate Statistics for the Social Sciences, 2nd Ed. Hillsdale, NJ: Lawrence Erlbaum Associates.

Wiggins, J.S. (1988). Revised Interpersonal Adjective Scales (IAS-R). Psychological Assessment Resources. Odessa, FL.

Combat and Non-Combat: Should Individual Values Differ?

Herbert George Baker, Ph.D.
United States International University

Abstract

In research with U. S. Marines, officers and staff noncommissioned officers were asked about the values they felt should be shown by junior and senior enlisted members in both combat and non-combat situations. Instrumentation was the SYMLOG questionnaire. Results indicate that there is little difference in the desired values across junior and senior members, or across combat and non-combat situations. Values profiles are also compared with an empirical teamwork values norm.

A perennial topic of discussion among U. S. Marines has been whether or not there is, or should be, a difference between the "combat Marine" and the "barracks Marine"; i.e., between the attitudes and behaviors shown in combat and non-combat situations. Anecdotal evidence abounds on both sides of the controversy. That the two situations can impose vastly differing requirements on the individual Marine is, of course, unarguable. However, should the values held by the individuals also differ? This research adds a quantitative dimension to the discussion.

Method

Subjects

Participating in the study were 81 officers and staff noncommissioned officers stationed at the Marine Corps Recruit Depot (MCRD) in San Diego. Subjects were selected based on availability. A requirement for participation was that the Marine had combat experience (defined as having drawn hostile fire pay).

Instrumentation

The SYMLOG system (SYstematic Multiple Level Observation of Groups) (Bales, 1988) measures individual and/or group values along 26 vectors, producing a values profile that can be compared with others or with a statistical norm profile. The scores also result in location along three orthogonal dimensions: Values on Friendly vs. Unfriendly Behavior, Values on Accepting vs. Opposing the Task-Orientation of Established Authority, and Values on Dominance vs. Submissiveness. The statistical norm for effective teamwork values is based on more than one million questionnaire administrations across the spectrum of occupations, age groups, geographic regions, and gender.

SYMLOG questionnaires use an introductory "context" or focus paragraph to orient the subject's thinking. Up to four questions are then posed, each question requiring response to 26 value statements. Here, the context paragraph and four questions were:

Focus: the culture of your organization

Think about your experience of your military organization in both combat and non-combat situations. Consider the way the members of your organization interact with each other. Reflect on the philosophy, policies, and procedures of your organization as these are played out on a daily basis over time. Reflect also on what is required in order for your organization to be successful and effective in accomplishing its mission. Keep these reflections in mind as you answer thw questions below.

Question 1: In general, what kinds of values need to be shown by junior enlisted personnel in order for your organization to be successful and effective in a COMBAT environment?

Question 2: In general, what kinds of values need to be shown by junior enlisted personnel in order for your organization to be successful and effective in a NON-COMBAT environment?

Question 3: In general, what kinds of values need to be shown by senior enlisted personnel in order for your organization to be successful and effective in a COMBAT environment?

Question 4: In general, what kinds of values need to be shown by senior enlisted personnel in order for your organization to be successful and effective in a NON-COMBAT environment?

The Field Diagram is a two-dimensional chart representing three dimensional group space. The vertical dimension represents Accepting/Opposing the Task Orientation of Established Authority, whereas the horizontal dimension represents Friendliness vs Unfriendliness. The Dominance/Submissiveness element (the third dimension) is represented by the size of circles denoting each individual or group, larger circles indicating increasing dominance. The diagonal double-pointed arrow (vector PF), represents a pathway, in effect, toward effective teamwork.

The Bargraph depicts mean scores for the group on each of the 26 SYMLOG questions. There is also a line connecting a series of Es, representing the statistical norm. Averaged scores on each question produce a bar composed of a string of Xs. Distance of the terminal X from the E represents the deviation of the group mean from the norm, shown visually and tested for significance.

Procedures

Data were collected at MCRD during Spring, 1995. Data collection was coordinated, and questionnaires distributed, through the office of the Chief of Staff. In effect, selection was random. Participation was voluntary; there was no time limit.

Results

Figure 1 portrays the mean responses for the four questions. The circles marked SCO and SNC represent Senior Enlisted, Combat and Senior Enlisted, Non-Combat, respectively. Similarly, circles marked JCO and JNC represent Junior Enlisted Combat and Non-Combat. Distance and size differentials among circles represent differences across groups and situations.

Also shown in Figure 1 is the circle (marked MEP) representing the position and size corresponding to the statistical norm for MOST EFFECTIVE. Deviation of participant mean responses from the statistical norm are apparent in location and size differences between the corresponding circles. Significant differences among circles are: JCO/JNC on UD (p=.05); JCO/SNC on UD (p=.01); and JNC/ SCO on PN (p=.01); all four differ from the MEP at p=.01.

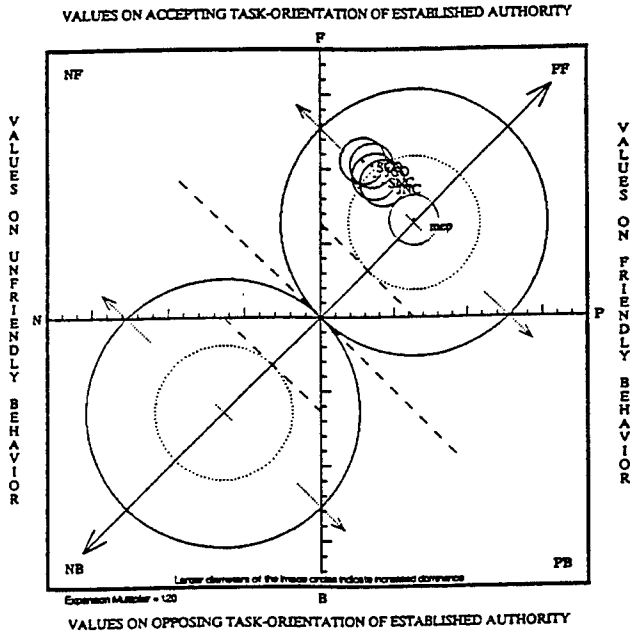


Figure 1. Group Average Field Diagram

Figures 2, 3, 4, and 5 show the bargraphs relating to the mean scores on each question. The scores are connected to produce a group profile on the 26 value questions. Also indicated on the bargraphs is the statistical norm for each question, marked by Es connected by a line, depicting a statistical norm profile for effective teamwork.

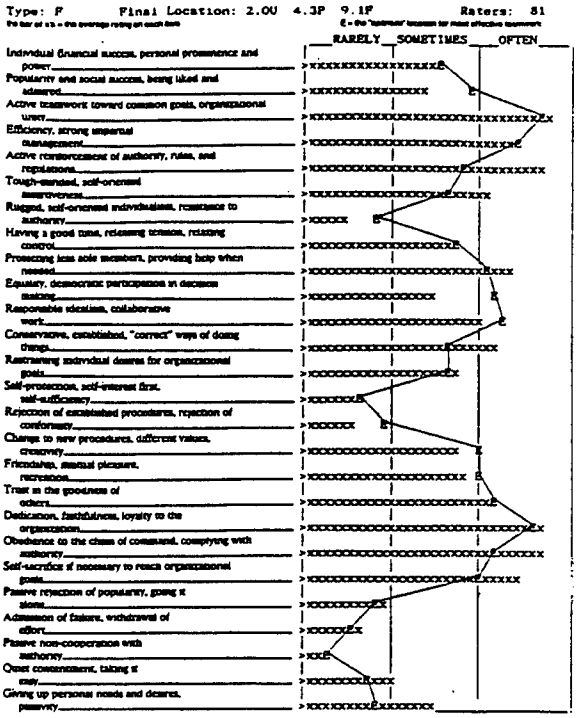


Figure 2. Bargraph
Junior Enlisted - Non-Combat

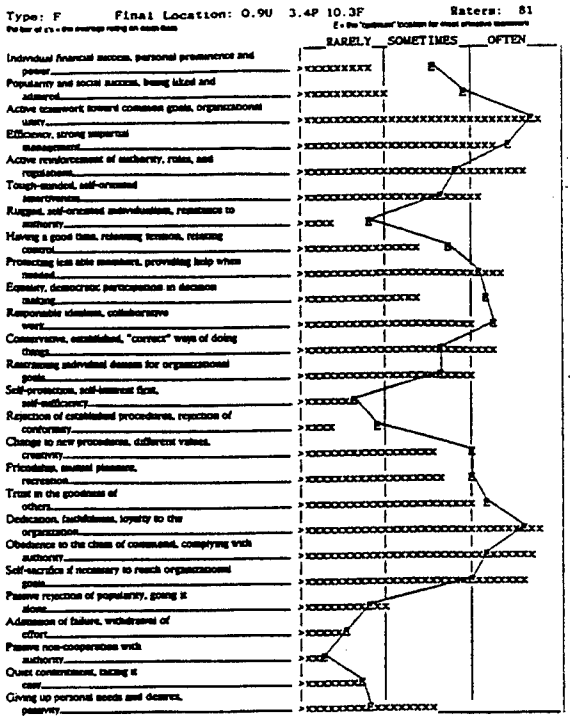


Figure 3. Bargraph
Junior Enlisted - Combat

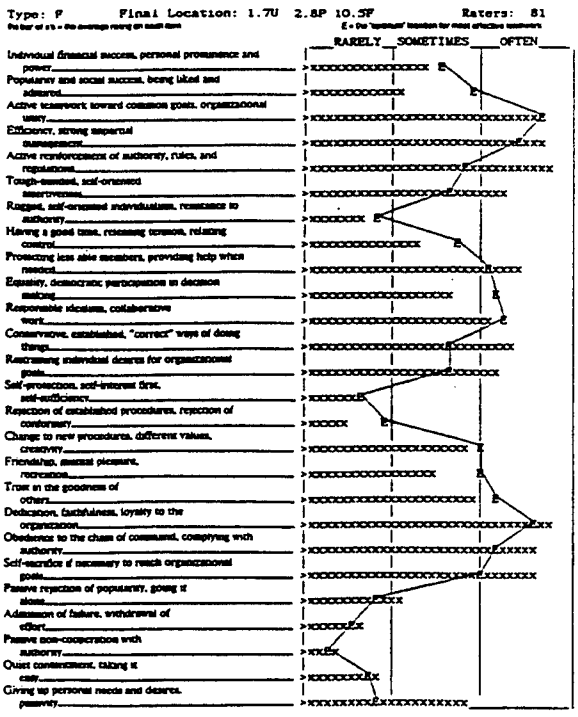


Figure 4. Bargraph
Senior Enlisted - Combat

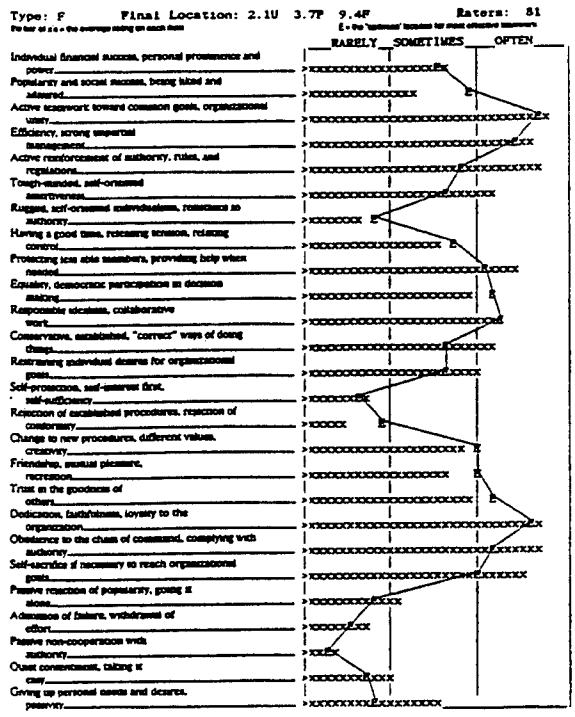


Figure 5. Bargraph
Senior Enlisted - Non-Combat

Differences between the senior and junior enlisted, combat and non-combat profiles and the norm profile are visible in the shape of the line and differences in the magnitude of the X bars (participants' mean scores) and the E points (normative points). Practically, differences of less than five Xs are non-significant. T-tests (two-tailed) for the significance of difference between two means were performed for each of the 26 questions. Results of those tests showed the following statistically significant (p=.05) differences between the relevant means:

SCO & SNC	Diff. on 0 Items	JCO & JNC	Diff. on 4 Items
SNC & JNC	Diff. on 2 Items	SCO & JCO	Diff. on 4 Items

Discussion

There is very close proximity among the four circles representing the two groups and two conditions in Figure 1. It is obvious that desired values in combat, for both groups, tend somewhat more toward acceptance of the task orientation of established authority, away from friendly behavior. The desired values for senior enlisted, in both combat and non-combat situations, show similar directional tendencies. All four circles distance themselves somewhat from the teamwork norm (MEP) in precisely the same direction. However, sizes of all circles (denoting dominance) are rather similar.

These findings indicate that Marine officers and staff noncommissioned officers view the values which should be held by junior and senior enlisted Marines as almost identical (only two differences in non-combat, four in combat). And, nearly all of the values shown in a non-combat environment are equally applicable during combat (no differences for senior enlisted, four for junior).

These results show that, while movement to a combat environment may impose serious change and need for adaptation, the values orientations of Marines will remain highly similar to those which lead to effectiveness in the non-combat environment. Also, training transfer will be facilitated by similar values expectations. Marines train for combat, and the values inculcated in training will be supported and supportive to a great degree during combat.

A number of values which junior and senior enlisted should show in both combat and non-combat situations differ from the statistical norm for effective teamwork (JNC, 15 items; SNC, 18 items; SCO, 17 items; JCO, 19 items). However, for practical significance, differences were far fewer: SNC, five items; SCO, seven items; JNC, seven items; JCO, nine items. Thus, there are some differences in teamwork values operative in combat, but, on the whole, teamwork values which are effective in peacetime prove largely effective in wartime.

References

Bales, R. F. (1988) A new overview of the SYMLOG system: Measuring and changing behavior in groups. In R. B. Polley, A. P. Hare, and P. J. Stone (Eds.), The SYMLOG practitioner (pp. 319-344). New York: Praeger.

The Relationship Between Environmental Attitudes and Environmental Behaviors Among Air Combat Command Members

Captain Daniel T. Holt, M.S.
Auburn University
Lt Col Steven T. Lofgren, Ph.D.
Guy Shane, Ph.D.
Major Kevin L. Lawson, Ph.D.
Air Force Institute of Technology

Abstract

Air Combat Command members were surveyed to determine the extent to which they held pro-environmental attitudes and how frequently they engaged in specific behaviors that were deemed environmentally protective. Results indicate relatively strong support for environmental issues, relatively infrequent environmentally protective behavior, and a moderate positive relationship between environmental attitudes and behaviors.

In an effort to mitigate the environmental effects of Air Force activities, the Air Force has focused its attention and its fiscal resources in four main arenas: restoration, compliance, conservation, and pollution prevention. Largely, these programs have been directed towards problems specific to the workplace. Recently, one component of the Air Force which manages 29 bases, Air Combat Command (ACC), has attempted to expand its recycling programs, composting programs, and hazardous materials collection programs to include individual activities outside of the workplace. Consequently, the organization's leaders have recognized the need to foster individual commitment in order for these programs to be successful and meet their objectives.

As these programs continue evolve, ACC hopes to foster this commitment through the integration of pro-environmental attitudes and pro-environmental behaviors into everyday life. This has brought us face-to-face with the classic problem of the attitude-behavior relationship. This study was designed to determine the extent to which ACC members held pro-environmental attitudes and how frequently they engaged in specific behaviors that were deemed environmentally protective. Additionally, it determined if there was a correlation between an individual's attitude toward the environment and their behavior.

Many researchers have assessed the extent to which different groups hold pro-environmental attitudes (e.g., Arcury, 1990; Noe and Snow, 1990). Generally, they have suggested that most citizens hold deep-seated pro-environmental attitudes. However, these studies did not provide any empirical data to indicate whether individuals that subscribe to the pro-environmental attitudes measured engage in more ecologically responsible behaviors.

Still, many presume that those who have a higher or deeper level of concern for the environment are more likely to act in an ecologically responsible manner. Thus, many researchers

have attempted to measure the statistical correlation between an individual's environmental attitude and his or her environmental behavior (e.g., Van Liere and Dunlap, 1981; Scott and Willits, 1994). The results consistently indicate a weak positive correlation between an individual's environmental attitude and his or her environmental behavior.

Although we assumed that environmental attitudes among ACC members reflect those of American society at large, this particular relationship has not been investigated. Moreover, we believe it is relevant to determine how unique segments of the population differ with regard to environmental attitudes and behavior.

Method

Environmental Attitudes

Environmental attitudes were measured using twelve items devised by Dunlap and Van Liere (1978). Each of the items was accompanied by five response categories: (1) Strongly disagree, (2) Mildly disagree, (3) No Opinion, (4) Mildly Agree, and (5) Strongly Agree. The ratings were collapsed into three categories labeled Disagree (by combining the mildly and strongly disagree selections), No Opinion, and Agree (by combining the mildly and strongly agree selections). In addition, the final four items on the survey were negatively phrased and reverse scored.

While Dunlap and Van Liere suggest that the twelve attitude items measure a single environmental attitude, researchers have found that these items may measure up to three separate attitudes (e.g., Albrecht et al, 1982; Scott and Willits, 1994). From data collected during a pilot study conducted at Wright-Patterson AFB, factor analysis, using varimax rotation, suggested a three factor solution was appropriate. Cronbach's alpha (ranging from .77 to 0.81) suggested that each of the attitude factors had sufficient reliability to warrant use.

Environmental Behaviors

Environmental behaviors were assessed using eleven items that were hypothesized to measure two principal behaviors. Each of the items was accompanied by the following scale of five responses: (1) Always, (2) Most of the time, (3) Occasionally, (4) Seldom, and (5) Never. The ratings were collapsed into three categories labeled Never/Seldom (by combining the never and seldom selections), Occasionally, and Usually (by combining the most of the time and always). Factor analysis, using varimax rotation, suggested the two factor solution was appropriate. Cronbach's alpha (0.85 and 0.84) suggested that each of the behavior factors had sufficient reliability to warrant use.

Air Combat Command Data

A total of 312 ACC members returned completed questionnaires. Members were randomly selected based upon social security number and ranged in grade from enlisted to officer.

Summary statistics were used to determine the extent to which members showed support for environmental issues and participated in environmentally protective behaviors. The bivariate correlation among the factors was calculated. This technique determined if a member's expression of support for environmental issues was related to the frequency that the member participated in environmentally protective behavior.

Results

Environmental Attitudes

Generally, ACC members indicated support for the pro-environmental position expressed by each of the attitude items (see Table 1). The data suggest that ACC members believed that man was abusing the environment, and his interference with nature often leads to disastrous consequences. These results are displayed in Table 1.

Balance of Nature. Over 75% of those queried agreed that the balance of nature is delicate and easily upset while a much higher percentage (nearly 85%) agree that humans must live in harmony with nature in order to ensure human survival. In addition, nearly 80% of the members believed that mankind is severely abusing the environment. Fewer (70%) believed that when humans interfere with nature it often produces disastrous consequences.

Limits to Growth. The results dealing with this factor were more inconsistent. While the members generally indicated a mild level of agreement with the concept of the earth's and the economy's limits, a few items seemed to vary considerably. Nearly 66% of respondents agreed with the idea that the earth has limited resources and room. Yet, less than half indicated that they thought there were limits to the growth of industrialized society. In contrast, slightly more than half of the members agreed with the notion that there are limits to economic growth and economic growth must be controlled (55%).

Man Over Nature. Overall, the data indicated that the members disagree with the notion nature exists merely as a resource for human exploitation. The majority of individuals rejected the belief that humans were created to rule over nature. Similarly, the majority (57.5%) of respondents rejected the idea that humans have the right to modify nature to suit their needs. In addition, nearly three quarters disagreed with the statement, humans need not adapt to the natural environment because they can remake it to suit their needs.

Table 1: Environmental Attitudes of Air Force Members

ITEM/FACTOR	Percent Response		
	Disagree	No Opinion	Agree
BALANCE OF NATURE			
The balance of nature is very delicate and easily upset.	20.2	3.5	76.3
When humans interfere with nature, it often produces disastrous consequences.	23.8	7.1	69.2
Humans must live in harmony with nature in order to survive.	9.3	7.7	83.0
Mankind is severely abusing the environment.	13.8	7.4	78.8
LIMITS TO GROWTH			
We are approaching the limit of the number of people the earth can support.	25.3	22.8	65.8
The earth is like a spaceship with only limited room and resources.	16.0	14.4	69.6
There are limits to growth beyond which our industrialized society cannot expand.	24.7	26.3	48.0
To maintain a healthy economy we will have to develop a steady state economy where industrialized growth is controlled.	21.6	22.6	55.8
MAN OVER NATURE			
Mankind was created to rule over nature.	53.6	14.4	31.9
Humans have the right to modify the natural environment to suit their needs.	57.5	9.9	32.7
Plants and animals exist primarily to be used by humans.	63.2	10.3	26.5
Humans need not adapt to the natural environment because they can remake it to suit their needs.	74.8	12.9	12.4

Environmental Behaviors

The majority of members indicated at least occasional participation in consumer/household environmentally protective behavior. However, the majority indicated less than occasional participation in environmentally related social activities. These results are displayed in Table 2.

Consumer/Household Practices. The demonstrated commitment to environmentally protective behavior within this subscale was inconsistent. Specifically, over 50 % of those questioned never/seldom avoid buying a product because it is not recyclable. Yet, nearly 39.4% usually avoid buying or using aerosol sprays. While 53.8% of the respondents never/seldom avoided buying products if they are not recyclable, it appeared that the majority of respondents

voluntarily recycle certain items on a regular basis (63.4% reported usual recycling of newspapers, glass aluminum, motor oil, etc.). Additionally, over 70% usually take more care in the use of chemicals.

Social Behavior. Overall, few members participate in environmentally protective social behavior. In all cases, nearly 60% of the members polled feel into the least frequent participation class for the activities identified..

Table 2: Environmental Behavior of Air Force Members

ITEM/FACTOR	Percent Response		
	Never/Seldom	Occasionally	Usually
CONSUMER/HOUSEHOLD PRACTICES			
Avoid buying or using aerosol sprays.	32.0	28.2	39.8
Specifically avoid buying a product because it was not recyclable.	53.8	31.8	14.4
Read labels on products to see if the contents are environmentally safe.	43.6	30.1	26.3
Use biodegradable plastic garbage bags, soaps, and other items.	30.5	28.5	41.0
Voluntarily recycle newspapers, glass, aluminum, motor oil, other items.	16.1	20.5	66.4
Take more care in the use of chemicals.	10.0	15.6	74.4
SOCIAL BEHAVIOR			
Boycott a company's products because of its record on the environment.	67.0	19.5	13.5
Contribute money to an environmental, conservation, or wildlife preservation group.	57.1	24.3	18.6
Attend a meeting related to ecology.	90.4	8.3	1.3
Do volunteer work for an environmental, conservation or wildlife preservation group.	83.3	13.1	3.6
Track my congressman's and senator's voting records on environmental issues.	82.0	12.2	5.8

Environmental Attitude - Behavior Relationship

The bivariate correlations among the five scores were all positive and statistically significant at the 0.0001 level (shown in Table 3). The balance of nature subscale had the largest r values linking it to the two behavior subscales. Specifically, the balance of nature subscale was linked to the consumer/household practices subscale with an r value of slightly more than 0.4. This suggests that when an individual believes in the notion that nature is a delicate, interdependent system that the same individual would more frequently participate in consumer/household practices that are considered to be environmentally protective. Similarly, an

individual having that same belief could be expected to more frequently participate in environmentally protective social behaviors (r value of 0.3). While these correlations suggest a positive relationship between the balance of nature subscale and the behavior subscales, these correlations are only moderate.

None of the other attitude-behavior correlations exceeded 0.28. This result suggests that there is a positive relationship between pro-environmental attitudes and participation in pro-environmental behavior. However, it also suggests that pro-environmental attitudes are not strong predictors of pro-environmental behavior.

Table 3: Correlations Relating Attitude Subscales and Behavior Subscales

	Consumer/ Household Practices	Social Behavior
Balance of Nature	0.40076	0.29682
Limits to Growth	0.23398	0.25136
Man Over Nature	0.28350	0.20327

Discussion

In conclusion, for environmentally protective actions to take place, pro-environmental attitudes and beliefs are necessary but may not be sufficient, given possible barriers and perceptions toward pro-environmental actions. Generally, ACC members express relatively strong support for environmental issues. However, they only occasionally engage in activities that contribute to the preservation or protection of the environment. This result suggests that additional environmental awareness programs are not needed. Instead, programs designed to influence participation and increase involvement in more prevention-oriented behaviors would prove to be more useful. Our results suggest that organizational leaders should explore alternatives to eliminate any barriers that prevent or discourage individuals from participating in environmentally protective behaviors.

References

Albrecht, D., Bultena, G., Hoiberg, E., & Nowak, P. (1982). The New Environmental Paradigm Scale. The Journal of Environmental Education, 13, 39-43.

Arcury, T. (1990). Environmental Attitude and Environmental Knowledge. Human Organization, 49, 300-304.

Dunlap, R. & Van Liere, K. (1978). The New Environmental Paradigm. The Journal of Environmental Education, 9, 10-20.

Noe, F. & Snow, R. (1990). The New Environmental Paradigm and Further Scale Analysis. The Journal of Environmental Education, 8, 20-26.

Scott, D. & Willits, F. (1994). Environmental Attitudes and Behavior: A Pennsylvania Survey. Environment and Behavior, 26, 239-260.

Van Liere, K. & Dunlap, R. (1981). Environmental Concern: Does it Make and Difference, How is it Measured. Environment and Behavior, 13, 651-676.

Yetton, P.W., & Johnston, K.D. (1992). Performance heteroscedasticity: Methodological threat or theoretical opportunity. Working Paper 92-04, University of New South Wales, Sydney, Australia.

Estimating the Utility of Organizational Change Using Probability-Based Simulations

Winston Bennett, Jr., Ph.D. and Robert M. Yadrick, Ph.D.
Armstrong Laboratory Human Resources Directorate
Bruce Perrin, Ph.D.
McDonnell Douglas Training Systems

Abstract

We used a probability-based simulation to determine the potential impact of a proposed organizational reengineering change and to (a) quantify requirements for additional training as a result of new personnel assignments; and (b) estimate the requirements for additional training, providing measures of training utility, changes in proficiency, and the impact of new training requirements on available training resources. We present the results of the simulation and discuss the implications of our results for identifying and quantifying the key factors associated with process and organizational structure reengineering efforts and for assessing the utility of change.

Organizational change activities typically are focused on refining or reengineering organizational processes and/or structure at one or more levels. However, it has been quite difficult to demonstrate that interventions have resulted in any systematic and quantifiable effects on the organization, beyond those associated with eliminating jobs and reducing the number of personnel. This is most likely due to the fact that the levels within an organization (e.g., an individual, workgroup, or division) tend to moderate our ability to determine the impact of the intervention or change. Given recent reductions in manpower, personnel, and training (MPT) and the perceived need to "rightsize" the Air Force to meet current defense demands, finding effective ways to assess the impact of interventions and reengineering activities becomes especially critical.

Conventional approaches to assessing the value-added or benefit of human resources interventions and organizational change have attempted to relate the intervention directly to measures of organizational productivity in a traditional utility analysis approach (e.g., Cascio, 1989). In general, utility analysis uses an estimate of the validity of a personnel intervention, such as a personnel selection method or a training program, and translates this parameter into organizational productivity in terms of dollars. However, certain aspects of implementing utility approaches have presented problems (e.g., Greer & Cascio, 1987). In addition, utility analysis may be quite limited as a method of determining the benefit one might actually expect from a training program if the effects of suspected moderators cannot be conceptualized easily with respect to the training program (Cascio, 1989).

For example, an organization may be reengineered to provide a different job structure in order to provide a more appropriate or flexible response to current demands. Changing the flows of personnel between jobs under this new structure invalidates the previous training utility analysis. The trainees' job experiences may be different in significant ways, calling into question the validity estimate used to evaluate utility and making a new validation study necessary to assure the applicability of the existing training program. Similar arguments could be made for organizational

reengineering ranging from the introduction of new materials and methods to the organization's size or typical span of control. We believe probability-based organizational simulation provides a more flexible method for assessing intervention utility under a variety of types of organization reengineering. The ability of training utility analysis to reflect organizational impact will be further eroded if Cascio's (1995) predictions about the fluidity of work in the next generation of organizations are realized.

An alternative approach to the assessment of the impact of change involves the use of a probability-based, organizational simulation technology such as the Training Impacts Decision System, or TIDES; see Vaughan & Yadrick, 1992; Mitchell, Yadrick, & Bennett, 1993. TIDES was developed by the U.S. Air Force for assessing the effects of reengineering activities. The simulation has the analytic capability to (a) provide a measure of training utility, expressed as changes in overall training costs and changes in the requirements for qualified personnel to support a new organizational structure; (b) quantify the reduction in requirements for additional training in some areas as a result of new personnel assignments; and (c) estimate the requirements for additional training in other areas. The TIDES model relates micro-level personnel events to macro-level outcome variables. Data for TIDES comes from a variety of sources, including job analysis, existing MPT data bases, and subject-matter experts= judgments. A computer simulation provides information on the flow of individuals through jobs, task performance requirements, and various formal and informal (e.g., on-the-job) training requirements. From these individual events, the system estimates task-level, on-the-job training events. Finally, the system estimates overall training resource requirements, costs, and capacities from the task-level events. Once the computer simulation of the current flows (the baseline) has been developed, plausible alternative flows and job and training structuring can be developed and new simulations can be conducted. Results from the alternative simulation outcomes can be compared to the baseline to examine impacts associated with the alternatives.

Method

The Air Force recently mandated a change to the way maintenance activities occur within operational Aerospace Propulsion units. The existing maintenance organizational structure was known as "3-level maintenance". The first level involves on-aircraft on the flightline, while the second, or intermediate, level involves work conducted in intermediate shops and the third involves shipping parts to and from the depot to the various operational units on the base. Parts that require repair or repaired components awaiting redistribution to the units are maintained at this level.

Procedure

The proposed restructuring of the Air Force maintenance organization involves the elimination of the flightline shops as a separate level. The changes would affect every aircraft maintenance occupation and would involve substantial changes to both the process of Air Force maintenance and the structure of maintenance organizations. The rationale for this change is that force drawdowns and fewer aircraft allows substantial savings in costs associated with MPT and logistics. In addition, it is expected that readiness, as measured by individual technician proficiency, will remain about the same. Figure 1 depicts the Aerospace Propulsion Job and Training structure for the 3-

level maintenance organization and the alternative 2-level structure. In the 3-level organizational structure, jobs 2 and 4-9 are performed by technicians in intermediate maintenance shops. Restructuring would eliminate these jobs and personnel would be reassigned to non-intermediate shop jobs. Although space limitations preclude a detailed discussion of the mechanics of the process here, a full description of this simulation, including the essential assumptions, is available from the first author.

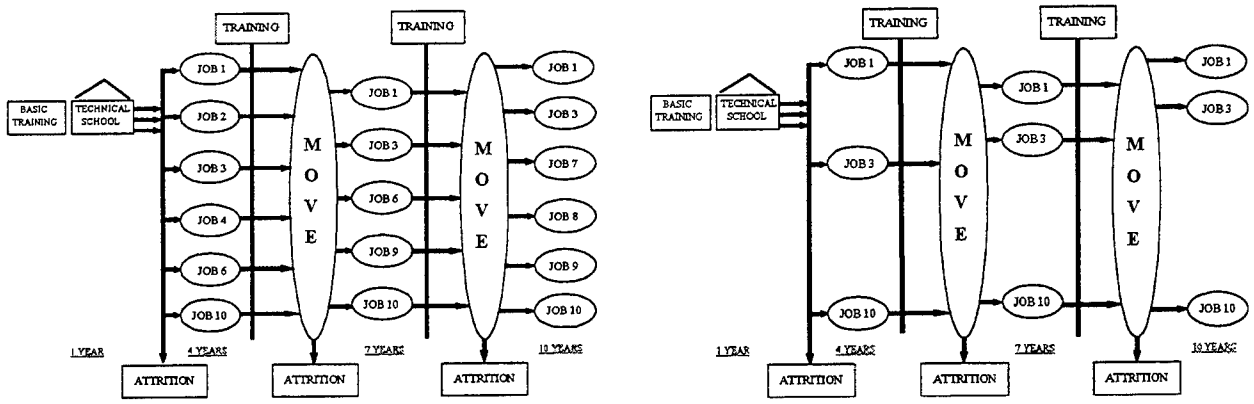


Figure 1. Baseline 3-Level and Restructured 2-level Maintenance Organizations

Thus, personnel who would previously have been assigned to the intermediate Aerospace Propulsion jobs are now assigned to jobs related to the new 2-level organization (that is, jobs 1, 3, and 10). One unexpected benefit of preliminary simulations is the suggestion that new reassignment policies may be needed because there will not be sufficient personnel to fill some 2-level jobs after 10-years, if present transition probabilities into these jobs are maintained.

Results

Figure 2 shows the results of our simulation and the impact of changing from 3-level to 2-level maintenance on overall training costs for the Aerospace Propulsion occupation. As the figure shows, training costs increase initially because of the additional duties that new occupation members assume at the beginning of their Air Force careers and the additional training burden that this imposes initially.

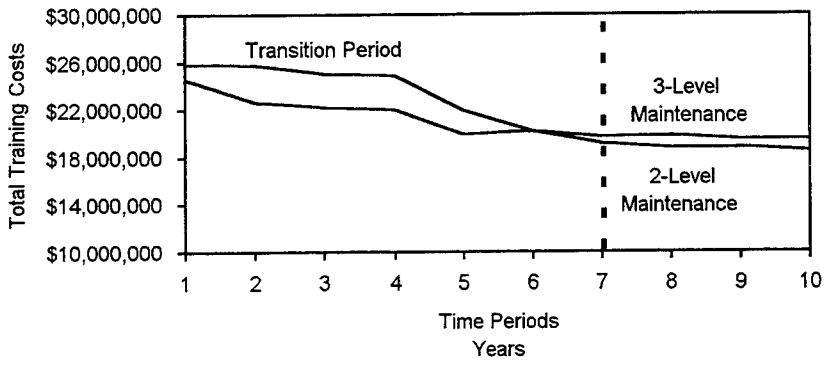


Figure 2. Estimated Training Costs for 3-Level and 2-Level Maintenance Organization.

The gap between the training costs under a 2-level structure compared to a 3-level structure begins to close after about five years, and after the nine-year point costs are lower for the new organization. Analysis of details of the simulation shows that this effect is due to (a) a reduction in subsequent training required for airmen later in their careers as a result of their receiving additional initial training and additional job experience throughout the first several years of their career, and (b) an overall reduction in personnel requirements associated with the new organization, which are reflected in Figure 3. This pattern continues through the 10-year point and shows no sign of reversing after that.

Figure 3 shows the simulation results of changing from 3-level to 2-level maintenance on the number of qualified airmen required to support maintenance activities in the Aerospace Propulsion occupation. As shown in this figure, there is a reduction in the number of qualified airmen required to support the 2-level maintenance organization for Aerospace Propulsion. There are several reasons for the observed reductions. As proposed in the original rationale for the restructuring, there would be a general decrease in the number of aircraft and operational flying wings as the defense requirements for the United States change over the next 10 years. This reduction was thought to potentially be due to the diminished nature of threats to U.S. interests around the world, which would reduce the need to maintain current levels of equipment resources and manpower in the Air Force. The results shown in Figure 3 capture the manpower reductions associated with the restructuring quite well.

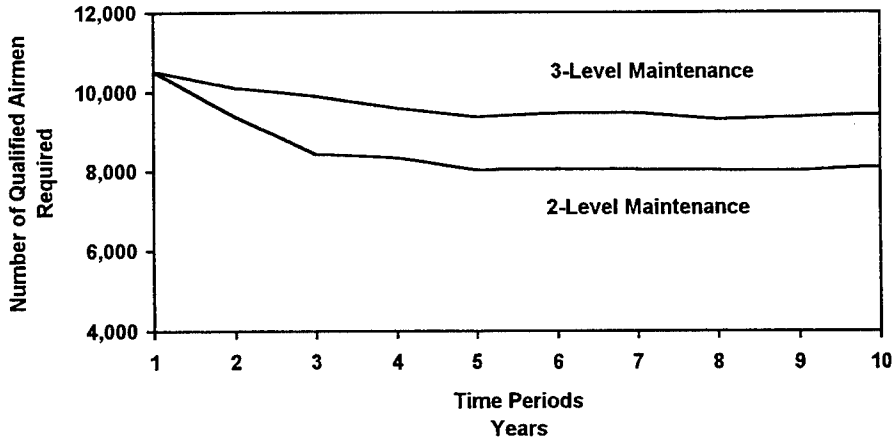


Figure 3. Number of Qualified Airman Required for Each Structure

It is also important to note that if the 3-level maintenance structure remained in place, all other factors held constant, the demand for qualified airmen to support the Aerospace Propulsion occupation would also remain fairly constant. In addition, it should be noted that the reduction in qualified airmen available under the 2-level is related to the overall training costs shown in Figure 3. That is, as the requirement for qualified airmen is reduced, the number of individual who would be sent to initial skills training and follow-on craftsman training would likewise be reduced, thereby decreasing overall training costs.

Discussion

The results from the simulation provided information related to the potential reduction in requirements for additional training in some areas as a result of new personnel assignments. Further, these results are expressed as changes in overall training costs and changes in the requirements for qualified personnel to support a new organizational structure. Thus, they are relevant metrics for assessing the cost/benefit or utility of organizational change for the Air Force.

This simulation was developed at the request of senior AF managers, who wanted to identify and quantify the effects of the proposed change to the maintenance structure. The results show that this is feasible, and also demonstrated that the restructuring would result in substantial training cost and personnel saving. However, the simulation also suggests that these savings would not be realized until after a transition period of several years.

Also, the simulation provided some surprising additional information related to assignment policies. Recommendations for restructuring, based in part on simulation analyses such as these, were used by the AF in decision making regarding this dramatic change to the maintenance community. It is important to note that the results from the simulation are based on changes that affect a single Air Force occupation. In reality, a substantial number of career fields are affected by the 3-level to 2-level maintenance restructuring. Therefore, it is conceivable that the results obtained in our simulations of the Aerospace Propulsion occupation would be similar for other occupations.

Author's note: We thank Brice Stone and Kathryn Turner of Metrica, Inc. for running the simulations and summarizing the results discussed in this paper.

References

Cascio, W. F. (1989). Using utility analysis to assess training outcomes. In I. L. Goldstein (Ed.). Training and Development in Organizations. San Francisco, CA: Jossey-Bass.

Cascio, W. F. (1995). Whither industrial and organizational psychology in a changing world of work? American Psychologist, 50 (11), 938-939.

Greer, O. L. & Cascio, W. F. (1987). Is cost accounting the answer? Comparison of two behaviorally based methods for estimating the standard deviation of job performance in dollars with a cost-accounting-based approach. Journal of Applied Psychology, 72, 588-595.

Mitchell, J. L., Yadrick, R. M., & Bennett, W., Jr. (1993). Estimating training requirements from job and training pattern simulations. Military Psychology, 5, 1-20.

Vaughan, D. S., & Yadrick, R. M. (1992). An organizational analysis simulation technology. In W.E. Alley (Chair). Organizational analysis issues in the military. Symposium presented at the annual conference of the International Military Testing Association, San Diego, CA.

Computer Adaptation of Task-based Occupational Analysis to the Changing World of Work

William J. Phalen, M.Ed.
Jimmy L. Mitchell, Ph.D.
The Institute for Job and Occupational Analysis

Abstract

Task-based occupational analysis can be readily adapted to the study of process- and team-based organization of work activities, even in the present rapidly changing world of work. The importance of task data is enhanced when valid and reliable estimates of absolute time spent on tasks are obtained through automated procedures. Tasks are too fundamental to the way work is organized and perceived by workers to be laid aside in favor of abstract descriptors.

In a comprehensive and forward-looking paper on the nature and implications of the dramatically changing world of work, Cascio (1995) notes "the growing disappearance of 'the job' as a fixed bundle of tasks" (p. 930) He observes that work requirements for both workers and managers are beginning to exhibit a growing emphasis on core competencies that are "virtual, boundary-less, and flexible" (p. 930) to meet demands of customers and threats of competitors that are constantly changing. Cascio detects a shift away from a task-based toward a process-based organization of work that lends itself to the formation of autonomous work groups or process teams of varying size and duration. He states that workers today have to engage in continuous learning in order to adapt to changing circumstances and should be prepared for multiple careers. Thus, he asks:

> What will be the future of traditional task-based descriptions of jobs and job activities? Should other types of descriptors replace task statements that describe what a worker does, to what or whom, why and how? Will 'task cluster' statements or 'subprocess' statements become the basic building blocks for describing work? What does a job description look like in a process-based organization or work? (p. 932).

This paper shows why and how task-based job analysis remains relevant even in the ever changing work environments described by Cascio.

Flexible Approach to Task-based Job Analysis

If we begin with the notion that a task is the smallest unit of work that a worker normally uses to define what he or she does in the workplace, it would appear that tasks possess some very desirable measurement properties: (a) they represent well-defined homogeneous chunks of work that a worker can comprehend and reliably rate on one or more unidimensional scales; (b) they can be used as movable and replaceable components for defining or designing any job, process, or

team effort at any moment in time. If changes occur in the composition of a job, process, or team, these changes will be most clearly detectable in terms of the removal, insertion, revision, or replacement of tasks.

Cascio (1995) suggests that dimensions of work other than tasks may become more important descriptors of work in the future -- such as environmental, contextual, social, and personal dimensions, in addition to the more traditional knowledge, skill, and ability dimensions (p. 932). However, we must first define the work requirements in terms of specific tasks in order to accurately assess the applicability of these dimensions. Otherwise, we are relying on general perceptions and hunches. To the extent that various dimensions can be linked to specific tasks or task clusters, work can be restructured to accommodate work requirements to the types of personnel available. Rather than considering dimensions such as abilities and interests, for example, as non-task dimensions, they should be treated as characteristics of tasks (or task clusters). This is accomplished by merging workers' biodata with their task response data from occupational surveys. Some of the biodata is obtained from a background section included in the occupational survey. Other biodata, such as aptitude or academic variables, is extracted from personnel files and merged with the survey biodata. It will then be possible to obtain average dimensional values for each task. For example, you might obtain the average pay grade level for each task based on the pay grade levels of those who perform the task. Similarly, you might compute the average interest or aptitude level for each task. Likewise, the percentage of workers who perform each task and use a given knowledge or tool may be computed, if these have been included as background items in the occupational survey. Tasks can then be clustered, not only on co-performance, but also on their profiles across a defined set of dimensions, such as knowledge, skills, and abilities (KSAs), to arrive at clusters of tasks with similar profiles. Tasks in the same co-performance cluster might be assigned as a functionally homogeneous unit of work. Tasks in the same KSA cluster represent feasible cross-training options or structural components in a restructured work environment.

In some instances, it may be desirable to obtain dimensional data on tasks directly, rather than through the cross-multiplication of biodata with task data. Important dimensions, such as task criticality, as measured by "consequences of inadequate performance," is best obtained for each task by having an appropriate number of subject matter experts provide ratings (as determined by interrater agreement criteria). If obtaining task-level ratings on a given dimension seems to be too labor intensive, such ratings might be obtained on task clusters instead, such as the co-performance or KSA-based task clusters described above. In order to support a process- or team-based approach for organizing work, the background section of the occupational survey section will have to include variables that identify the team(s) the workers belong to and/or the process steps or subprocesses the worker is associated with. Thus, task-level data can be aggregated for a team or a process, and work relationships of workers assigned to a team or process can be analyzed and, if desired, be realigned according to worker characteristics identified in the biodata and associated task characteristics.

The ability to cluster tasks into meaningful clusters at higher and higher levels of aggregation to meet the needs of various levels of users highlights the flexibility of a system which obtains data at the most specific level that is feasible. Small chunks of task-level data can be

aggregated to any level of generality that might be useful, but data gathered at a less specific level cannot be disaggregated to answer questions requiring more specific data.

Automation of Occupational Analysis

If task-based occupational analysis as described above seems too cumbersome, too labor intensive, too static, recent developments in the Air Force's Comprehensive Occupational Analysis Programs (CODAP) software system have done much to alleviate such complaints. First of all, programs have been developed which automate much of the analysis process, including the selection and interpretation of significant job types and task clusters from a hierarchical clustering of jobs or tasks. The core tasks and discriminating tasks within the selected clusters are identified for further analysis and reporting to manpower, personnel, and training managers. Secondly, procedures have been developed for the electronic distribution of occupational surveys to personal computers (PCs) worldwide at a moment's notice, as well as the PC-based capturing and transmission of survey response data to a central computer. The PC-based, self-administration procedure allows tailored presentation of background and task items to raters (workers or subject matter experts) using probabilistic branching techniques to limit the number of items that need be presented. Feedback mechanisms have been incorporated to prompt the rater to evaluate and correct "suspicious" responses identified by algorithms embedded in the survey software. Thus, large amounts of task-level data obtained from thousands of workers can be analyzed with quick turnaround and relatively little administrative overhead.

The automation of the process of distributing occupational surveys and the electronic capturing of response data allows the task-based approach to react rapidly to the changing world of work. Thus, it will be feasible to take frequent "snapshots" of the world of work, either on a periodic basis, such as a worker's birthday (continuous saturation sampling), or as a specific need arises (focused sampling). Specific needs may also require updated survey instruments. Rapid revision of survey instruments will, of course, be no problem in an automated environment. It is evident that the computerization of the entire survey development, distribution, response capturing, and analysis process has converted the task-based occupational analysis process from a cumbersome dinosaur to a dynamic, interactive, flexible process capable of keeping pace with a dramatically changing world of work.

One area in which the PC has made an important contribution is in the application of complex scaling procedures for rating tasks. In particular, the PC has made it possible to obtain from workers estimates of absolute time spent on tasks that are more valid and reliable than those derived from four competing scales: a relative time spent scale, a direct magnitude estimation scale, an indirect magnitude estimation scale, and an end-anchored graphical scale. Descriptions of the scales and research findings are reported in Albert et al. (1995) and Phalen (1995).

The Absolute Time Spent (ATS) Scale

The amount of time a worker spends performing a task is a complex concept composed of two less complex components that are more psychologically manageable: frequency of task performance during a specified period of time and the amount of uninterrupted time it normally takes the worker to perform the task once. Frequency estimation has proved to be especially accurate and reliable. The estimation of time is more subject to influences in the rater's internal and external environments; however, the estimation of time for a single performance of a task is a well-defined event of limited scope. Once accurate and reliable estimates of frequency and time to perform a task once have been obtained, the total amount of time spent within a specified period is nothing more than the cross-product of the two component measures rescaled to a common metric.

The measurement of total absolute time spent (ATS) and its component subscales of "frequency of task performance" and "time to perform a task once" are an integral part of the Air Force's PC-based Computer-Administered Survey Software (CASS) system (Albert, et al., 1995). All estimates of frequency and time are provided by the rater in natural language form by selecting codes and inserting numeric values. While previous approaches used to estimate ATS have been plagued with problems of overestimation, the CASS system has been able to incorporate a number of operationally tested features that seem to have largely overcome this problem (Phalen, 1995).

Overall, the CASS system has been found to be easy to use and reliable (average coefficient for individual raters over a two-to-four-week period = .66). Also, raters have selected ATS scale estimates as significantly more valid than those of the four alternative scales (a value of $p < .001$ was associated with most of the computed Chi-square values). Upon completion of a survey administration, the total absolute time spent vector, as well as its subscale vectors, are immediately available on floppy disk as a data file.

From an organizational analysis standpoint, there is much to be gained from the information provided by the component subscales of the ATS estimation procedure, as follows:

(1) The "frequency of task performance" subscale could be used as a measure of the need for refresher training. Infrequently performed tasks that have high hazard potential or serious consequences if performed inadequately may require occasional refresher training. On the other hand, the occurrence of mishaps and accidents may be related to either low or high frequency of performance, and this may vary from task to task.

(2) The "time to perform a task once" subscale also has numerous applications that should be of interest to organizational analysts. The average amount of time it takes various functional subgroups of workers to perform specific tasks at various grade or experience levels could be used to set standards for these groups and the various levels within groups. On the other hand, if certain individuals within a group are requiring much more time or much less time to perform these tasks, chances are that the high-time workers may need training or motivation, while the low-time workers may either not be doing the task as it should be done or have valuable time-

saving expertise that should be tapped. The average length of time it takes entry-level personnel to perform specific tasks, together with the associated standard deviations, might be used to set bypass criteria and standards for OJT or formal training.

(3) Significant differences in task performance times between selected subgroups might indicate that the task in question is not really the same task for the various subgroups. This could occur, for example, if the subgroups represent equipment operators or repairmen on different aircraft types who rate many of the same task statements.

(4) The effectiveness of two different training environments could be compared by determining how long it takes the average worker trained in either environment to perform specific tasks soon after beginning the same entry-level job.

(5) Perishability of skills for specific tasks could be determined by computing the functional relationship between the frequency with which specific tasks are performed and the time it takes workers with similar background characteristics to perform these tasks.

(6) Work descriptions for individuals and groups would be a much richer source of information if the "frequency of task performance" and "time to perform a task once" data were shown together with the total (cross-product) ATS values. Clustering of work descriptions would provide clearer and more meaningful results if all three vectors of data were clustered as one profile, using a common metric, or as the average of three overlap matrices.

Conclusion

This paper has attempted to show that task-based occupational analysis can be readily adapted to process- and team-based organization of work, even in a rapidly changing world of work. It has also attempted to show the enhanced importance of task data when valid and reliable estimates of absolute time spent on tasks are obtainable through automated procedures. The "task" is much too basic to the way work is organized and thought about by workers to be laid aside in favor of less fundamental, more abstract descriptors.

References

Albert, W.G., Phalen, W.J., Selander, D.M., Dittmar, M.J., Tucker, D.L., Hand, D.K., Weissmuller, J.J., & Rouse, I.F. (1994). Large-scale laboratory test of occupational survey software and scaling procedures. Proceedings of the 36th Annual Conference of the International Military Testing Association (pp. 241-246). Rotterdam, The Netherlands: European Members of the IMTA.

Cascio, W.F. (1995). Whither industrial and organizational psychology in a changing world of work? American Psychologist, 50, 928-939.

Phalen, W.J. (1995). A critical evaluation of various procedures for estimating time spent. Proceedings of the 37th Annual Conference of the International Military Testing Association (In Press). Ottawa, Ontario, Canada; Canadian Forces Applied Research Unit.

Task-Based Analysis of Processes

Brice M. Stone, PhD.
Kathryn L. Turner, MBA
Metrica, Inc.
Robert C. Rue, PhD.
SRA Corporation
Sharilyn A. Thoreson, PhD.
Jimmy L. Mitchell, PhD.
McDonnell Douglas Training Systems

Abstract

Much of corporate America has embraced business process reengineering (BPR) to survive in today's competitive environment. This paper addresses a methodology for converting and using an existing career field data collection procedure in conjunction with an existing occupational data base to implement process reengineering. The Government Performance and Results Act of 1993 and the Department of Defense Corporate Information Management initiative have both provided motivation to improve business processes. DoD and the Air Force, in particular, have the advantage of being able to used a scientifically defendable and well used source of information, occupational analysis data, which could enhance and facilitate their process reengineering objectives. The Air Force occupational analysis program focuses on identifying discrete tasks which are clustered into work units to be performed by specialists. Refocusing the occupational analysis program to conform to processes which may or may not cross specialist lines will provide the opportunity to evaluate process reengineering alternatives with a richer set of criteria. The Occupational Measurement Squadron at Randolph Air Force Base in San Antonio, Texas maintains occupational data associated with each of the 200+ 5-digit Air Force specialties or career fields. In addition to the AFS tasks lists, OMSq maintains information concerning characteristics of each of the tasks such as task learning difficulty, percent members performing the task, training emphasis, etc. This information can be used to assess the effect of reengineering on training requirements, aptitude requirements, manning requirements, productivity, and other evaluation criteria.

Much of corporate America has embraced business process reengineering (BPR) to survive in today's competitive environment. In their book, Reengineering the Corporation (1993), Michael Hammer and James Champy define reengineering as:

> "the fundamental rethinking and radical redesign of business processes to achieve dramatic improvements in critical, contemporary measures of performance, such as cost, quality, service, and speed."

One legacy of modern business and the science that supports it is the division of work into simple, easily trained tasks or jobs, and the assignment of those tasks or jobs to specialties. Using this approach, the big picture -- satisfying the customer through delivery of quality products or services -- is sometimes lost. Specialists are each responsible for only their small portion of the work and lack the insight or motivation to adopt a larger view. Even managers do not always view the business as a process. When quality drops and customers are dissatisfied, managers often fail to consider the process; instead they focus on specific tasks in their search for origins of the problem. This focus limits the solutions they consider and the results they achieve.

Hammer and Champy (1993) identify "processes" as one of the key words in their definition. The old paradigm requires complex oversight processes to ensure that acceptable products result from the combined output of many simple tasks. It is these complex oversight processes that keep individual workers from understanding the big picture; they also provide managers inadequate control over the quality of the product. BPR seeks to replace complex processes with simple, flexible processes. Simplifying processes can also reduce costs by eliminating nonvalue-adding tasks and other inefficiencies.

Congress and the Department of Defense (DoD) have also embraced BPR. The Government Performance and Results Act (GPRA) of 1993 and the DoD Corporate Information Management (CAM) initiative have both provided motivation to improve business processes. The GPRA provides for the establishment of strategic planning and performance measurement in the Federal Government as techniques for improving the efficiency and effectiveness of government programs. In particular, a stated purpose of the GPRA is to "... improve Federal program effectiveness and public accountability by promoting a new focus on results, service quality, and customer satisfaction." The GPRA requires all Federal agencies to submit, by September 30, 1997, strategic plans for their program activities. These plans must define the agency's mission and state measurable goals and objectives for achieving their mission "... including a description of the operational processes, skills and technology, and the human capital, information, and other resources required to meet those goals and objectives."

CAM requires the Services to reengineer business processes of functional areas such as personnel and logistics before they are allowed to invest in new information technology. CAM uses facilitated subject matter expert (SME) workshops with representatives from the functional area to design the future, or TO-BE, activities that form the new business processes. Initiatives are identified to move the functional area closer to the reengineered processes. These initiatives often include the insertion of new information technology. Functional economic analysis is then used to compare initiatives and to develop a business case for deciding which initiative to select for funding. The business case typically requires that the initiative pay for itself over some planning horizon. The Air Force has been involved in a number of joint CAM efforts and has also initiated several of its own.

One of the results of BPR is to redefine the way work is organized. Hammer and Champy (1993) identify several recurring themes in the new processes: (1) Previously distinct jobs are combined into one, compressing processes horizontally by having the teams perform several, sequential tasks; (2) Processes are also compressed vertically by allowing the workers or teams of workers to make decisions that formerly were made by management -- decision making becomes part of the process; (3) The steps in the process are performed in a natural order, not in an order dictated by the

old, complex process, thus removing artificial precedence relationships and allowing more tasks to be performed simultaneously; (4) Processes are more flexible and less standardized; (5) More work is accomplished across organizational boundaries, reducing reliance on specialists; (6) More emphasis on cross-functional teams, sometimes called Integrated Product Teams (IPTs) which bring together specialists from several disciplines to produce a particular product; and (7) Nonvalue-added tasks are eliminated. This includes minimizing reconciliation and reducing checks and controls to only those that make economic sense. By sharing databases and reducing the number of data input points, the need for reconciling data is reduced.

Occupational Analysis Data

The Air Force occupational analysis program supports a traditional approach to performing business processes wherein discrete tasks are clustered into work units to be performed by specialists (Christal, 1974). As noted above, this approach often results in workers and managers who do not have the big picture of the product being produced or the service being performed. The ongoing occupational analysis program tends to make marginal changes to occupational clusters. Changes may indeed produce improvements in performance and use of resources, but the narrow focus on making minor changes to the existing AFS starting points produces only limited solutions to problems and potentially fails to address larger scope process improvements.

To fully reap the benefits of reengineering, the Air Force would need to modify or redesign its occupational analysis program to support the larger scope business processes. Reengineering seeks to simplify processes that have grown complex through evolution. Critical examination of processes often reveals components that are no longer producing added value or can be simplified through technology. By first examining and reengineering the underlying business processes, the rich data from occupational analysis can be more effectively used to structure jobs.

The methodology for this modification/redesign follows a similar approach to the present methodology used by the Air Force for collecting occupational analysis data. The development process begins with the identification of the task inventory list for the process. The Occupational Measurement Squadron (OMSq) at Randolph Air Force Base in San Antonio, Texas maintains task lists associated with each of the 200+ 5-digit Air Force specialties (AFSs) or career fields. In addition to the tasks lists, OMSq maintains information concerning characteristics of each of the tasks such as task learning difficulty (TD), percent members performing the task (PMP), relative percent time spent (PTS), training emphasis, etc. This information can be used to assess the effect of reengineering on requirements for training, aptitude, manning, etc.

AFS to Process Conversion

Technology has been developed, such as the Training Impact Decision System (TIDES), which demonstrates the ability to define jobs within career fields, as well as career fields, and training courses as a combination of tasks or task modules (Gosc, Mitchell, Knight, Stone, Reuter, Smith, Bennett, & Bennett, 1995). A task module (TM) is a group of tasks which are naturally performed or trained together in such a way as to take advantage of coperformance or co-training. These TMs are then

used to define the jobs within a career field. The manning requirements imposed upon these jobs, and the TMs of which they are comprised, form the basis for the demand for training on these TMs.

In the same way in which a career field is defined as a combination of jobs which are a collection of TMs, a process can also be defined in terms of jobs which must be performed and the TMs which comprise those jobs. For example, Hammer and Champy (1993) define a business process as a collection of activities (tasks or TMs) that takes one or more kinds of inputs and creates an output that is of value to the customer. The information compiled through the development of task lists can be redirected from career fields and career field training to processes and the jobs which are required to perform the processes.

Air Force training has always been oriented towards career fields; however, much of Army training is oriented towards units. For example, the nature of ground combat requires crews to be cross trained to fill in for each other and that equipment operators also be maintainers. The crew members are all cross trained for positions other than their own. Tank crew, howitzer crew, etc., perform first level maintenance on their equipment in the field. They accompany the equipment to second level maintenance and assist with the maintenance. Even ambulance crews are trained in vehicle maintenance. Thus, the Army defines training and jobs by units which perform specific tasks, task modules, or processes. For much of the Army, processes are already defined by a task list and Army occupational analysts can take advantage of the information which a well defined task list can render.

Air Force processes can be a combination of jobs which presently reside within several career fields (specialties) or within a single career field (specialty). If the jobs associated with the performance of a particular process all reside within the same career field, many of the advantages of a task (TM)-based approach to defining the process are minimized, e.g., the skill and knowledge requirements may be the same for the career field as for the process.

When the jobs which comprise a process are drawn from jobs across several career fields, some advantages can be gained in analyzing the reengineering of the process through a task (TM)-based approach. Tasks and TMs can be mapped to skill and knowledge requirements which can be used to identify training requirements (Moon, Driskill, Weissmuller, Strayer, Fisher, & Kirsh, 1991). Using tasks or TMs to define processes provides the basis for identifying skill and knowledge requirements and, thus, training needs which are directly tied to the process. Defining and constructing training courses based on process requirements may take advantage of a more natural, work-oriented order of performing tasks (TMs) and, thus, introduce larger economies of co-training and coperformance which are neglected or ignored when the focus is on career field requirements.

One of the keys to mapping TMs/tasks to processes will be to identify the TMs/tasks which comprise the process. Since this methodology does not exist, several alternatives will be discussed. One such alternative would be to assemble SMEs to identify, from a master TM/task list, those TMs/tasks which comprise the process. This is similar to the methodology which OMSq presently uses when updating or initiating a new AFS study, i.e., providing a task list to SMEs to identify the appropriate list of tasks active for a career field. The question is how to identify a beginning TM/task list which does not encompass the total task list across career fields.

One proposed methodology for accomplishing a reduction in the task list to a manageable level for review by SMEs is by using the Uniform Airman Report (UAR) to identify jobs and, thus, TMs/tasks associated with those jobs which are a part of performing the process. Several data elements from the UAR would be reviewed as candidates for assembling the original TM/task list. Data elements contained in the UAR such as functional account codes, location (organization/base/unit), job description, etc., could provide a basis for the identification of jobs or activities associated with processes. One or combination of these data elements will be reviewed to determine the best approach for identifying the original task list, and, thus, the TMs/tasks associated with the process.

Once the data element or combination of data elements have been used to identify individuals involved in the performance of the process, the individuals can then be mapped to Occupational Survey (OS) data. The tasks which these individuals have identified as the ones which they perform in the respective jobs will form the basis for the original process task list to be reviewed and refined by SMEs.

Once the process task list has been refined by SMEs, then the OS data provides an extensive amount of information which can be used to analyze process reengineering alternatives from numerous perspectives. For example, relative time spent performing tasks can be used to identify time intensive tasks which could be identified for process reengineering and/or technology improvements. Alternatives for restructuring of processes can be reviewed from numerous perspectives such as training requirements, knowledge and skill requirements, aptitude requirements, etc.

References

Christal, R.E. (1974). The United States Air Force Occupational Research Project. Lackland Air Force Base, TX: Occupational Research Division, Air Force Human Resources.

Gosc, R.L., Mitchell, J.L., Knight, J.R., Stone, B.M., Reuter, F.H., Smith, A.M., Bennett, T.M., & Bennett, W. (1995). Training Impact Decision System for Air Force Career Fields: TIDES Operational Guide. Brooks Air Force Base, TX: Human Resources Directorate, Technical Training Research Division, Armstrong Laboratory.

Hammer, M., & Champy, J. (1993). Reengineering the Corporation: A Manifesto for Business Revolution. New York: Harper Business.

Moon, R.A., Driskill, W.E., Weissmuller, J.J., Strayer, S.J., Fisher, G.P., & Kirsh, M. (1991). Using task co-performance modules to define job requirements. Proceedings of the 33rd Annual Conference of the Military Testing Association (pp. 243-252). San Antonio, TX: Armstrong Laboratory, Human Resources Directorate and the USAF Occupational Measurement Squadron.

Analysis of Outcomes for an Entity Based Job and Training Simulation Model

Kathryn Turner, MBA
Brice Stone, Ph.D.
Metrica, Inc.
Guy Curry, Ph.D.
Texas A&M University
Captain Teresa Bennett
Armstrong Laboratory/Human Resources Directorate

Abstract

This effort focused on examining a probabilistic entity based job and training simulation model, the Training Impact Decision System (TIDES), to develop a methodology for determining an optimum time period for stopping the time series simulation. The primary objective is to obtain and ensure representative outcomes from the simulation for baseline and alternative scenario simulations. Mathematical equations for the identification of the optimum time period for stopping the simulation were developed and will be imbedded into the TIDES modeling system. Confidence intervals were developed for outcome values of the simulation to allow comparison of scenario simulations.

This paper describes a research effort which focused on developing a methodology to determine the optimum time period for stopping a time-series entity based simulation. This work is being conducted using the Training Impact Decision System (TIDES), an entity based job and training simulation model. The TIDES is a probabilistic simulation in which entities, or airmen, are moved through a probabilistic horizon of jobs and training courses to simulate an airman's career. Data used in the simulation are from the TIDES study of the Electronics, Computers and Switching career field (2E2X1).

The primary objective of this sensitivity analysis was to obtain and ensure representative outcomes from the simulation for baseline and alternative scenario simulations. Outcome variables of interest included: total training costs, formal training costs, on-the-job (OJT) costs, and total force level. A methodology for determining when the simulation had reached steady-state was determined. Equations were also developed to determine the appropriate length of the simulation from steady-state to ensure representative outcomes. Equations for confidence intervals for outcome variables were also developed.

Method

This sensitivity analysis was composed of three primary steps. The first step was to identify when the simulation reached steady-state. The second step was to then compute the required simulation length after reaching steady-state. The final step was to calculate mean values for the outcome variables of interest and confidence intervals associated with those estimates.

The methodology to be used in this sensitivity analysis was originally presented in Stone, Turner, Curry & Bennett, 1995.

Several scenario simulations were used in the sensitivity analysis. The baseline scenario simulation was run with all data set to their default values. Several policy changes were then implemented within the TIDES system to create alternative scenario simulations. These alternative scenarios were created to ensure that the TIDES simulation was reaching steady-state in a similar time period across the baseline and alternative scenarios. The alternative scenarios created included deleting a job from the career field and deleting a training course from the career field.

The TIDES simulation was run for a period of 125 years for the baseline and each of the alternative scenarios. The TIDES simulation does not begin with an existing inventory in the career field at the start of the simulation period. Instead, TIDES allows the population to build over time in the simulation or "grows" the force. As a result, there is an initial bias in the simulation output which must be addressed. Figure 1 illustrates this initial bias in the simulation output for the outcome variable total training costs from the baseline scenario.

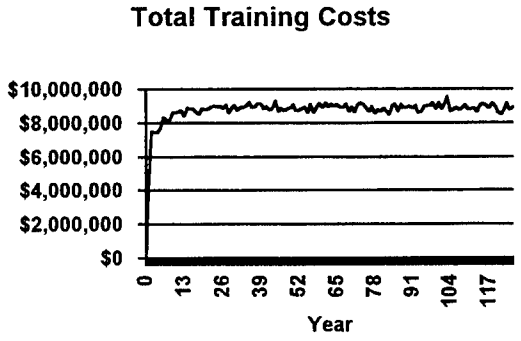


Figure 1. Total Training Costs for Baseline Scenario

These initial observations must be removed from the estimates of the outcome variables or these estimates will be biased. If it is assumed that the sample means of the start-up period converge to the steady-state mean, then a technique for dealing with the start-up period can be developed. Generally this is accomplished by deleting the first k observations from the simulation run, and then using the unbiased estimator of the population mean μ, which can be expressed as:

$$\overline{X}(n,k) = \frac{1}{n-k}\sum_{i=k+1}^{n} X_i \qquad (1)$$

where k is the number of observations deleted and n is the total number of observations from the run of the simulation (Curry, Deuermeyer & Feldman, 1989).

Welch (1983) suggests the simplest and most general technique for selecting the k number of observations to delete is a graphic procedure. Welch's procedure is to make N replication runs of the simulation, where N is at least greater than four. The replicated runs of the simulation are then averaged and then graphed against time. The value of k is selected where the mean value of all replications tends to "flatten out." Welch's procedure was applied to the baseline and alternative scenarios running each scenario five times. The results of the average total training cost of five runs for the baseline simulation is presented in Figure 2.

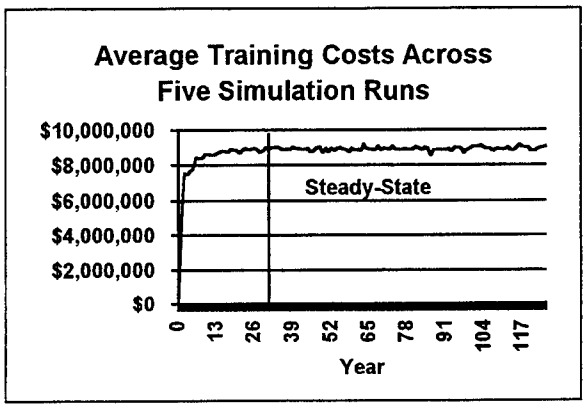


Figure 2. Average Results for Baseline Scenario

From Figure 2, the timing for reaching steady-state is somewhere in the vicinity of 30 years. By the 30 year point, the average of the five runs has "flattened out." Similar results were seen for other outcome variables for the baseline scenario, as well as in the results from the alternative scenario simulation runs. Therefore, the value which was used for k will be 30 years. By the 30 year point, the simulation has reached steady-state. It is reasonable to expect this to occur by the 30 year point since the simulation begins with a zero population and the average airman's length of service in the output is approximately ten years. Parameter estimates of the means and standard deviations were estimated for each outcome variable using (n-k) observations from a run of the simulation.

Now that the time period in which the simulation reaches steady-state has been identified, the next step was to develop parameter estimates and confidence intervals for the outcome variables and to determine the required length of the simulation. Mean values for each of the outcome variables can be estimated using equation 1, filtering out the initial start-up bias period. The variance was then calculated, filtering out the initial bias period, using the following equation:

$$S^2(n-k) = \frac{1}{n-k-1}\sum_{i=k+1}^{n}(X_i - \overline{X}(n-k))^2 \qquad (2)$$

These parameter estimates were then used to develop confidence intervals for each of the outcome variables. From these parameter estimates, the 1-α confidence limit (such as the 95% level of confidence) for the filtered mean estimated was computed using the following equation:

$$\overline{X} \pm t_{n-k-1,\alpha/2} \frac{S(n-k)}{\sqrt{n-k}} \tag{3}$$

where $t_{n-k-1,\alpha/2}$ was a critical value based on the Student's-t probability distribution with n-k-1 degrees of freedom (Curry, et al., 1989).

Equation 3 was also used in estimating the required simulation run length (for years beyond the start-up period) in a sequential process (Curry, et al., 1989). First, estimates of $\overline{X}(m)$ and $S^2(m)$ were obtained for an outcome variable, where m was some number of observations beyond the start-up period. The half-width confidence intervals for the filtered parameter estimates of the outcome variables were then examined to determine if they are within the desired accuracy. These half-width confidence intervals were calculated using the equation:

$$v = t_{m-1,\alpha/2} \frac{S(m)}{\sqrt{n}} \tag{4}$$

where v is the one-half width of the confidence interval. The simplest method to establish the required run length is to set a minimum run size, such as 10 observations beyond the start-up period, and run the simulation to this point. Using these observations, the mean and standard deviation can be computed and then the half-width confidence intervals computed using equation 4. If the half-width of the resulting confidence interval is within the desired accuracy then the process is terminated. Otherwise a fixed number of additional observations are simulated and the test is repeated using the larger sample size. This process is repeated until the estimates result in an acceptable error. If a very tight requirement is placed on the confidence interval, then the number of observations required can become very large. This methodology was used to determine the appropriate run lengths for scenarios in the TIDES simulation.

Results

The initial results from the sensitivity analysis showed that the simulation reached steady-state within the first 30 years. Therefore, the start-up bias period was assumed to occur in the first 30 years, and these were the k observations which were removed from all parameter estimates. The same start-up bias period was seen in all the simulation runs for the baseline and alternative scenarios.

For the baseline and alternative scenarios, parameter estimates were obtained for the outcome variables of total training costs, format training costs, OJT costs and force level. The simulation was initially run 10 years beyond the start-up point and half-width confidence intervals were calculated for each outcome variable. For this sensitivity analysis, it was assumed that if the half-width of the 95% confidence interval was less than 0.5% of the mean estimate, the estimate was sufficiently accurate. The initial run of 10 years did not produce sufficient parameter estimates to meet the accuracy criteria, so the simulation was run for an additional 10 years (total of 20 years plus the 30 year start-up period. The resulting parameter estimates for the baseline and alternative scenarios are shown in Table 1 for the outcome variable total training costs. The

table includes estimates of the mean, standard deviation, 95% confidence interval, half-width confidence interval and accuracy criteria (the half-width confidence interval percent of the estimate of the mean). As the table shows, all three scenarios passed the accuracy test when the simulation was run for 20 years beyond the start-up period. These parameter estimates would be reported from the TIDES software package and can be used to make comparisons of parameter estimates across different scenarios.

Table 1
Parameter Estimates for Total Training Costs Across All Scenarios

	Mean	Standard Deviation	95% Confidence Interval		Half-Width Confidence Interval	Accuracy Measure
Baseline	$8,950,982	$90,259	$8,907,642	$8,994,322	$43,340	0.48%
Delete Job	$8,690,770	$86,514	$8,649,229	$8,732,311	$41,541	0.48%
Delete Training Course	$8,887,293	$70,387	$8,853,495	$8,921,090	$33,798	0.38%

Discussion

The results of this analysis will be implemented within the prototype system of the TIDES. Simulations will always be run for a minimum of 30 years before outcome variable information is collected from the simulation results. Users of the system will be prompted for desired levels of accuracy of the outcome variables. For example, the user may specify that the confidence interval be no larger than 1% of the mean value for the outcome variable. The equations and methodology for determining run length presented in this paper will be automated within the TIDES system to establish for the user the required simulation length beyond the initial 30 year start-up period.

References

Curry, G., Deuermeyer, B., & Feldman, R. (1989). Discrete Simulation: Fundamentals and Microcomputer Support. Oakland, CA: Holden-Day, Inc.

Law, A. & Kelton, W. (1991). Simulation Modeling & Analysis. 2nd Edition. New York: McGraw-Hill, Inc.

Stone, B., Turner, K., Curry, G. & Bennett, T. (1995). Identifying Representativeness of Outcomes for an Entity Based Training Simulation Model. The Proceedings of the International Military Testing Association.

Welch, P. (1983). The statistical analysis of simulation results. The Computer Performance Handbook, 268-328.

Air Force Occupational Measurement Squadron Customer Satisfaction Survey Report: A Summary of Findings Regarding Examinee Knowledge of the Testing Portion of the Weighted Airman Promotion System[3]

1Lt Heather M. Henderleiter
Air Force Occupational Measurement Squadron

Abstract

This study was directed by the Air Force Occupational Measurement Squadron (AFOMS) Test Development Flight Quality Council to collect feedback from recent promotion test examinees. Pencil-and-paper surveys were administered to 163 examinees from 81 career fields and 5 major commands. These surveys served as starting points for focus group discussions of the issues surveyed. Through the discussions, we discovered areas of concern regarding an apparent lack of examinee knowledge, including what the WAPS Catalog is (the source for determining study references for examinees) and how to use it, how and when to challenge questions they think are faulty, the target range of means AFOMS uses when developing tests, and how to obtain references to study for their WAPS tests. This anecdotal evidence suggests a widespread problem within the enlisted corps that should be addressed. This apparent fundamental lack of knowledge could be corrected by the addition of a section in AFPAM 36-2241, Promotion Fitness Examination Study Guide, which is the study reference for the Promotion Fitness Examination (PFE).

The Weighted Airman Promotion System (WAPS) affects enlisted personnel in the grades of E-4 through E-6. The points each member earns for the six factors of WAPS (time in service, time in grade, awards and decorations, Enlisted Performance Reports (EPRs), Specialty Knowledge Test (SKT), and PFE) are added to reach a total score. The SKT and PFE each contributes 22% of the total score. Because these factors carry the highest weights, they tend to determine who will be promoted and who will not.

AFOMS brings in subject-matter experts (SMEs) from each career field to develop the tests for their specialties. These SMEs provide the technical expertise, and the psychologists assigned to our squadron provide the psychometric expertise. We make every effort to ensure that, as stated in the SME orientation briefing, "we provide our Air Force with the fairest, most valid, most credible means possible of selecting the most knowledgeable airmen for promotion." It is to this end that the Test Development Flight Quality Council decided to investigate how examinees felt about test booklet appearance, testing procedures, SKT item content, SKT references, and the PFE.

[3] The ideas presented in this paper do not necessarily reflect the views of AFOMS, the US Air Force, or the DoD.

Method

Subjects

We surveyed 163 recent examinees (114 E-5s and E-6s and 49 E-4s) from 11 bases. Personnel from Air Education and Training Command, Air Mobility Command, Air Combat Command, Air Force Materiel Command, and Air Force Special Operations Command participated in the survey. We obtained data from 81 Air Force Specialties (AFSs). Five respondents did not supply us with AFS information.

Procedures

We conducted 15 focus group sessions at 11 bases. Eight were conducted during the E-6/7 testing cycle (15 Jan - 31 Mar) and 7 were conducted during the E-5 testing cycle (1 Apr - 15 Jun). We used five pencil-and-paper surveys to stimulate discussion in the areas of interest. The areas were test booklet appearance, testing procedures, SKT item content, SKT references, and the PFE.

We arranged meeting areas (typically conference rooms or classrooms) through the base test control officers (TCOs). Each TCO sent us a list of examinees who were scheduled to test on two specific days prior to our visit. We invited examinees via letter to attend our focus group sessions. We also informed the supervisor of each examinee that his/her subordinate had been invited to attend.

Each focus group session followed essentially the same format. We introduced ourselves and told the respondents that not only did we want to get their opinions on the five areas dealing with WAPS testing, but that we were also there to answer questions they might have about testing issues. For each of the five survey areas, we first asked participants to rate each of the items on our questionnaire. We used the ratings as a springboard to begin a group discussion of that item. We allowed 15 minutes of discussion for each questionnaire. At the end of each session, we opened the floor to discuss anything that the respondents might not have brought up during the appropriate block, and we also answered any related questions that participants posed.

Results

A recurring issue throughout all focus group sessions was an apparent lack of examinee knowledge about WAPS testing. Areas where this fundamental lack of knowledge were most apparent were the procedures to challenge test questions, the WAPS Catalog and issues surrounding it, how to obtain references for WAPS testing purposes, and general testing knowledge.

Over 90% of examinees surveyed indicated it is important to them that testing procedures allow them to challenge test questions they think are faulty. Although examinees almost universally expressed approval of the theory of challenging test questions, very few had ever actually attempted to query a question. A majority of the comments examinees made were related

to ideas they had to improve the system. Most did not seem to be aware of the process (initiate the query in the test room, provide justification within five workdays, wait for reply from AFOMS through the TCO). For example, a respondent told us, "You need to allow more time to reference questions with materials to validate your point against that question." Examinees are briefly informed of the process just prior to taking the test as the proctor reads the instructions. Approximately five percent of the examinees we spoke with told us they had been told by their test proctor that they could only challenge a question if they could provide the justification in the test room at that time.

When we included questions regarding the WAPS Catalog in our survey, we expected that examinees would have at least heard of it. When examinees are notified of their test dates, they sign a statement which indicates that the examinee understands that it is his/her responsibility to look up the study references in the WAPS Catalog. We asked each focus group about their experiences with the catalog. The results are summarized in the table shown below.

Grade	% who have heard of catalog	% who used catalog
E-5/6	58	35
E-4	33	10

Examinees repeatedly told us they had never seen the WAPS Catalog; that they did not even know it existed. They seemed to rely on word of mouth to determine what references they needed to study. For example, in 1995, Chapter 13 of the PFE Study Guide was not testable. At the beginning of the E-5/6 testing cycle in January, a majority of examinees did not seem to know that the information had been available since August 1994 (when the WAPS Catalog was published). By the time we surveyed E-4s, most examinees had at least appeared to have heard the rumor that they did not need to study that particular chapter. Very few examinees told us that they found the information in the WAPS Catalog.

In addition to using the WAPS Catalog to determine study references, examinees also need to obtain those references. For most career fields, this was not difficult in 1995. This was the first year the Extension Course Institute (ECI) sent out a set of Career Development Courses (CDCs) to each member eligible for promotion. Some examinees who tested early in the testing cycle did not receive their CDCs at least 30 days prior to testing. However, most seemed to have had no problem obtaining CDC references. Again, examinees did not seem to know that they should check their study references against the WAPS Catalog to determine if they had received the correct volumes.

Some general testing information of which examinees did not seem to be aware includes the range of means AFOMS uses to evaluate tests, why it takes so long to receive scores, and when testing takes place. AFOMS strives to maintain the mean of each test between 48 and 60%. This results in examinee's scores being lower than they are used to seeing. Many examinees indicated that when they saw that they had only scored around 50% that they felt they failed. There is no pass or fail in WAPS tests. Test scores are added to the points earned from the other factors to arrive at a final score. It takes months to receive scores because the final results are not in until and all deletions are processed and the testing cycle is over. One examinee wrote, "I was

under the impression that testing (for E-5) did not begin until May." This comment highlights the lack of knowledge some examinees have regarding WAPS testing.

Discussion

In general, we found that examinees are lacking in knowledge. They do not know the procedures to challenge a question, how to obtain a list of study references, or what the target range of means for the tests is. In order to increase their knowledge of their promotion system, a section should be added to AFPAM 36-2241, Promotion Fitness Examination Study Guide. This section should detail information about WAPS testing. We discovered that examinees seem to be so concerned with the possibility of being accused of cheating that they do not get the information they need. For example, examinees told us they do not ask about references because they did not think they could talk to anyone about any facet of testing. They do not know how to obtain the WAPS Catalog -- or even what it is. In some organizations, the WAPS monitors are nonexistent; in others, uninformed.

The enlisted personnel to whom we spoke indicated almost universally that they appreciated that someone had asked them what they thought and had taken the time to tell them things they need to know to be effective test takers. This information should be readily available to all personnel, and adding a section to the PFE Study Guide would be an efficient and effective means of disseminating it.

Quality of Life in the United States Army Recruiting Command:

A Qualitative Research Study

H. Michael Hughes, Ph.D.
Lieutenant Colonel, U.S. Army
United States Military Academy

Abstract

In response to Congressional pressure and concerns of senior Army leaders at the highest levels, the author conducted a qualitative research study to investigate factors impacting on the quality of life of recruiters, family members and support staff assigned to United States Army Recruiting Command. During a four month study spanning the United States, structured interviews were conducted with a purposeful, representative sample of personnel assigned to recruiting duty. Analysis of the qualitative data and available quantitative data from Recruiting Command discovered seven interdependent "findings" or themes. These seven findings represented the factors which most significantly detract from positive quality of life for recruiters and their families. Uncoordinated and inadequate medical support, lack of spouse and family involvement, critical levels of stress, lack of resources and support for mission accomplishment at station level, short term leadership focus, lack of preparation for racial and cultural diversity issues in recruiting, and inadequate financial support (variable housing allowance) for families arose as the seven major factors. Recommendations for action by senior Army leadership are being staffed and implemented as a result of this study.

Previous studies on morale, job satisfaction, stress and other "quality of life" concerns for recruiters have provided only broad, quantitative indicators of problem areas for recruiters and family members. A recent Department of Defense (Department of Defense, 1994) study reported seventy four percent of recruiters believed their families were not sufficiently prepared for the stressors inherent in a recruiting assignment. This 1994 study, however, did not provide sufficient behavioral descriptions of the specific factors which negatively influence psychological and physiological health in Recruiting Command. Only experienced recruiters with over one year in the role of recruiter had been surveyed in 1994; a gap in data from the new recruiter existed. The link between quality of life and coping strategies and effective accomplishment of the recruiting mission had not been studied in a comprehensive manner. Thus, the purpose of this study was to answer the following research questions. First, what is the relationship between mission accomplishment and quality of life for recruiters? Second, what are the variables that constitute quality of life for recruiters and their families? Third, how effective is the system for selection, preparation and socialization of the recruiter family into the role of Army recruiter?

Method

Qualitative research methodology was utilized to gain an appreciation for the complexity, interdependence and behavioral nature of factors influencing quality of life. A structured interview protocol was developed to facilitate discussion of relevant recruiter concerns. Following guidance from the U.S. Army Deputy Chief of Staff for Personnel, the focus was on "real world" concerns of soldiers and their family members. As described below, a purposeful, representative sample was selected from throughout the Recruiting Command. The entire life cycle of a recruiting assignment was examined from selection for duty, training and preparation, socialization into the role of recruiter at unit level, actual performance, evaluation and support.

With full cooperation of the U.S. Army Recruiting Command (USAREC) Commanding General and staff, the study began with an initial examination of records, files and internal documents. This provided insight into the various potential areas for investigation. Consultation with the Personnel Staff Directorate led to nomination and selection of specific units, sites and organizations for interview sessions. Once the interviews had been conducted, the researcher also served as a participant observer in the USAREC Family Life Symposium. During this forum, initial trends and themes were validated and discussed in depth with recruiters, family member delegates to the conference, and USAREC staff members responsible for quality of life issues. Site visits were also conducted at station, company, battalion and USAREC level to gather more information on emerging trends.

Once these themes and key factors/issues were identified, the researcher carefully addressed these various findings with relevant officials at various levels throughout the Department of Defense. Care was taken to ensure prolonged engagement with USAREC and persistent observation of recruiter problem areas. Multiple sources and methods of data collection were utilized in an effort to achieve triangulation. A methodological log was kept and officers experienced in qualitative research methodology and USAREC were asked for consultation throughout the process. The research goal was to obtain credible, relevant findings leading to specific recommendationss.

Subjects

Interviews were conducted with recruiters, commanders, family members and support personnel throughout the USAREC command. All 5 brigade areas were in the continental United States. Sites were chosen to selectively represent urban, rural and suburban units. Units demonstrating high, medium and low mission performance (as defined by achievement of recruiting mission goals) were selected for interviews. Also, units that are located near (within 30 miles), medium distance (30-75 miles) and remote stations (over 75 miles from any military support installation) were visited. Representatives from all 41 battalions met with the investigator. Additionally, the investigator met in small focus group sessions with the Family Service Coordinators from all 5 brigades and 41 battalions. Interviews were also conducted with 5 health care professionals responsible for treatment of recruiters and their dependents. Interviews were also conducted with 4 previous commanders and spouses of USAREC company

and battalion size units. A total of 151 interviews of approximately one hour duration were conducted during the period June through October 1995.

Procedure

After selection of a unit for a site visit, the researcher would coordinate visit dates and specific locations with the unit and station commanders. Since the USAREC Commanding General had granted access to units and personally supported the research study, cooperation of units was not an issue. After a briefing on the purposes of the study and confidentiality of responses, subjects completed demographic sheets and necessary informed consent forms.

During the private interviews, the researcher followed a standardized interview protocol addressing the following areas: (1) Duty descriptions of both assigned and implied duties, (2) A typical work day, week and month description to gain a sense of job demands (3) Perception of success or failure in the role (4) Personal and/or professional goal attainment in this job (5) Description from the interviewee's perspective of overall personal quality of life in this assignment; description of family members' quality of life; comparison with other Army assignments (6) Identification of key factors/forces impacting on the recruiter and/or family member quality of life (7) Discussion of specific preparation (or lack of preparation) in various life areas for this assignment (8) Discussion of coping strategies utilized by the recruiter and family (9) Ideas or recommendations on actions the Army or any organization/leader could take to improve mission accomplishment and quality of life.

Specific use of open ended probes was used to obtain in depth information about these various issues. During individual and group sessions, the researcher would carefully ask about any organizational policies or procedures that either hindered or facilitated mission accomplishment and/or quality of life. At the conclusion of each interview and group session, the researcher thanked each subject and restated the promise to keep individuals and units anonymous during report preparation. The subjects were provided with the researcher's name, phone number and electronic mail address to facilitate further exchange of relevant information.

Results

The content analysis of the data revealed findings relevant to quality of life. First, the interdependence of psychological and physiological health of the recruiter and mission accomplishment was noted. These phenomena were clearly linked and not separate entities. Simply stated, the recruiter needed to have a meaningful life experience in order to recruit effectively; mission accomplishment led to both extrinsic and intrinsic rewards which facilitated higher quality of life. Second, a number of variables were identified in seven themes or trends that constitute the quality of life for the recruiter and family members. These are discussed below. Third, several problem areas were identified (included in these seven findings) with the complex system which identified, selects, prepares and socializes the soldier (and family) into the role of Army recruiter.

These seven findings and themes were:

(1) Medical support to active duty recruiters and their family members in locations away from military posts is uncoordinated and woefully inadequate. This is the number one concern of recruiters and family members and has serious psychological, physiological, and financial impacts on the members of the command.

(2) Spouses and family members throughout USAREC feel a lack of involvement in the recruiter's role and in the unit. This has led to a feeling of alienation and a lack of family support for rigorous demands of the recruiting job of the soldier.

(3) Stress is a critical problem at the present time in specific units in USAREC. Cardiac arrests, stress related illnesses, divorce, child abuse and domestic violence related to job demands are at dangerous levels in various units. The chain of command magnifies this problem in certain cases due to overemphasis on mission accomplishment "at all costs."

(4) At station level, USAREC has not provided adequate support resources such as fax machines, coiers, and automation capability to facilitate mission accomplishment. Recruiters must work more demanding schedules and sacrifice personal time to cope with poor support.

(5) Leadership at station level is concentrated primarily on mission at the expense of quality of life and the long term physical and psychological health of the recruiters. Each month is treated as a production crisis and high levels of burnout are readily observable.

(6) Recruiters are not trained and prepared for racial, cultural and ethnic challenges in certain market areas. This hinders the establishment of trust between recruiter and prospect which negatively impacts mission accomplishment.

(7) The current finance system of Variable Housing Allowance calculation does not provide adequate support to USAREC soldiers assigned to high cost areas. This also negatively impacts mission accomplishment because long commutes and logistical problems lead to less time spent on actual recruiting tasks.

Discussion

For each of these findings, specific recommendations were made to address the problem areas. These recommendations were systemic in nature and required action from senior leaders in USAREC, Department of the Army and the Department of Defense. During the presentation, these recommendations will be discussed in detail.

This researcher found some recruiters to be literally "living on the edge" of the stress curve. The interviews produced a consistent, repetitive theme of recruiters facing the monthly pressure of mission tightening around their lives (and those of their families) like a band of steel which exerts more pressure as the end of the month approached. The phrase "36 one month tours" was heard repeatedly from recruiters, commanders and family members. The medical

concerns were real and urgent; having medical care denied or unavailable to family members caused incredible disruptions in recruiters' lives. As this researcher discussed this finding at the highest levels of USAREC, the senior leaders' frustration with the entire Department of Defense medical system was apparent.

As a result of this research study, the author developed a stress management workshop designed specifically for recruiters at station level. It was presented in a pilot workshop in late October 1995 and well received by the soldiers and families. It is currently being adopted in various formats for application throughout the command.

The researcher acknowledges the limitations of the study. This study was obviously conducted during a brief four month period on a purposeful yet limited sample and was able to obtain only a limited "snapshot" of the quality of life equation for recruiters. While the qualitative, structured interview protocol approach was useful in gaining a rich, "thick description" of factors related to quality of life, it was also time intensive. Due to funding limitations, the researcher collected data independently without the assistance of another team member. As a trained counselor, the researcher is aware of the clinical tendency to listen for language and nonverbal cues indicating depression and other psychological states.

Finally, the author wishes to thank the Deputy Chief of Staff for Personnel, United States Army, for funding support for this study. Also, the entire USAREC Command Group's cooperation and openness made this study successful.

References

Department of Defense. (1994). 1994 DOD Recruiter Survey. Washington, DC: Department of Defense Press.

Ability Of Military Recruits: 1950 To 1994 And Beyond

Brian K. Waters, Ph.D.
Human Resources Research Organization
Dana H. Lindsley, Ph.D.
Office of the Assistant Secretary of Defense (Force Management Policy)

Abstract

Changing ability of incoming recruits is a major measure of recruiting success. This paper examines civilian and military test score trends over the past four decades, using Scholastic Assessment Test (SAT), National Assessment of Educational Progress (NAEP), and Armed Forces Qualification Test (AFQT) score trends. Analyses are made by gender and race/ethnicity of military enlistment-age examinees.

The Department of Defense (DoD) monitors trends in the national population and within the Military Services to assure that the quality and quantity of the force meet operational needs. This paper looks at one aspect of national and military-age youth: measured cognitive ability. Four aspects of ability testing are addressed: 1) reference of scores to scores attained by a large group of examinees on the test ("norms"); 2) development of military norms since World War II; 3) test score trends for two major civilian testing programs and military enlistment testing during the past four decades; and, 4) the impact of the trends for military personnel selection and classification policy in the coming decade.

Test Norms

Scores on an ability or knowledge test convey little meaning in and of themselves. Reporting, for example, that a person correctly answered 70 percent of the test items is virtually meaningless unless one knows the appropriateness, content difficulty, and readability of the test for the examinees. A college senior correctly answering 70 percent would be quite different from a fourth grader performing similarly on that test. Thus, tests must be referenced (normed) to a comparable group's performance on them. The attribute measured is clarified if the test score is reported in relative terms, such as "scored in the 70th percentile of high school seniors nationally." Large samples of examinees' scores are used to set the relative score scales (norms) for civilian testing programs such as the Scholastic Assessment Test (SAT) or the National Assessment of Educational Progress (NAEP), and for military tests such as the Armed Services Vocational Aptitude Battery (ASVAB).

Military Selection Test Norms

The military (active duty and Reserve/Guard) tests over 1,000,000 applicants for enlistment annually. The scores are referenced to populations of examinees who took the test earlier. Scores from the Armed Forces Qualification Test (AFQT), a composite score comprised

of verbal and quantitative ASVAB test scores, have been used for enlisted applicant selection since 1950.

The 1944 Reference Population. The 1944 reference population included all males on active duty on December 31, 1944 (both enlisted and officers). It provided the basis of the military selection test system from the early 1950s through 1979, after which a new, more representative normative base, the 1980 Profile of American Youth, became the ASVAB reference population.

Profile of American Youth. The population of military-eligible youth had changed significantly between 1944 and the late 1970s. For example, education levels of our national youth had increased greatly, and demographic changes also had occurred in terms of race/ethnic proportions. It was questioned whether the 1944 reference population might still be representative of youth abilities over three decades later. At the time, the Department of Labor (DoL) was preparing to conduct a large-scale study of American youth from the ages of 14 though 21. By joining forces with DoL, DoD could gather nationally representative data for its new norms. Thus, in conjunction with the DoL's 1979 National Longitudinal Study of Youth Labor Force Behavior (NLSY79), DoD sponsored the PAY 80 study to establish a new normative base for military testing. After the NLSY79 survey data were collected from nearly 12,000 men and women 18 to 23 years old, the ASVAB was administered to the members of the DoL sample. Since the PAY 80 norming, major changes in the demographic characteristics of American youth have occurred, as shown in Table 1. The proportions of Hispanic and Asian youth have more than doubled, the relative proportion of Black remained unchanged, while White youth shrank by about 10 percent. The percent of "Other" minorities may have also diminished, however, the drop more likely reflects the category of "none" in the PAY 80 data, which was not present in the 1994 census.

Table 1. Composition of the U. S. Population by Race/Ethnic Group, 1980: (18 - 23 Years Old)[1] and 1994: (18 - 24 Years Old). (Numbers in Thousands)

Race/Ethnic Group	1980[2]		1994	
	Number	Percent	Number	Percent
White	17,784	73.3	17,177	68.7
Black	3,413	14.1	3,559	14.2
Hispanic	1,515	6.3	3,254	13.0
Asian	257	1.1	803	3.2
Other	1,271[3]	5.2	220[3]	0.9
Totals	24,240	100.0%	25,013	100.0%

[1] Excludes skips and refusals, and respondents responding "don't know" or "none".
[2] Estimated from Profile of American Youth data.
[3] Includes Native Americans, Alaskan Natives, Pacific Islanders, and respondents identifying themselves as "other".

Changes in geographic distribution, education level, and other variables likely related to measured abilities of military applicants have also occurred. As in 1980, DoD was able to co-sponsor a new NLSY project with DoL, the NLSY97. The PAY 97 study, begun in January,

1995, will gather computerized adaptive testing ASVAB and DoD-developed interest inventory, the Interest Finder, scores starting in early Summer 1997. The study will produce a new set of norms based upon the 997 American population of 18 to 23 year old youth. It will provide a scientifically sound basis for military selection and classification testing well into the 21st century. This brief review of norms development for military selection and classification testing over the past four decades provides the foundation for an analysis of trends in ability measurement since the early 1960s. The purpose of this discussion is to highlight the context of the abilities of current applicants to the Services

Four Decades Of Cognitive Test Score Trends

Large-scale, group-administered ability measurement had its roots in World War I military enlistment testing. National testing programs aimed at determining intelligence, aptitude, knowledge, skills, interests, and other domains of measurement have flourished since World War II. By tracking score trends on such tests, analysts can estimate the effects of changes in populations through the years. Knowledge of such trends has an important impact on educational, social, military, political, and economic policies. For the purposes of this paper, two major national civilian testing programs are shown: the Scholastic Assessment Test (SAT) and the National Assessment of Educational Progress (NAEP). Armed Forces Qualification Test (AFQT) score trends of military applicants are also discussed.

Scholastic Assessment Test (SAT) Score Trends

Table 2. Mean Scholastic Assessment Test (SAT) Verbal Scores by Gender - Biennial, 1972 through 1994.

	Verbal			Mathematics		
Year	Men	Women	Total	Men	Women	Total
1972	454	452	453	505	461	484
1974	447	442	444	501	459	480
1976	433	430	431	497	446	472
1978	433	425	429	494	444	468
1980	428	420	424	491	443	466
1982	431	421	426	493	443	467
1984	433	420	426	495	449	471
1986	437	426	431	501	451	475
1988	435	422	428	498	455	476
1990	429	419	424	499	455	476
1992	428	419	423	499	456	476
1994	425	421	423	501	460	479

The SAT comprises tests of verbal and mathematical abilities widely used for college entrance screening. It is scaled to a 1941 standardization population. Approximately 1,500,000 high school seniors take the SAT annually. It should be noted that SAT scores are <u>not</u> nationally representative, despite the large sample size. Only college-bound seniors' scores are included in SAT data. Table 2 displays SAT-Verbal score trends by gender from 1972 through 1994. Over

the 23 year period, SAT-Verbal scores showed a steady decline through the early 1980s, a slight increase in scores in the mid '80s, and a "flattening out" since the early '90s. Male SAT-Verbal scores dropped 29 points over the period, while female Verbal scores slipped, similarly, an average of 31 points. Table 2 also shows that SAT-Mathematics scores declined through 1980, followed by gradual increases until current scores in 1994. Over the 23 years, female SAT-Math scores dropped a single point, while male scores dropped 4 points. Table 3 displays changes in SAT-Verbal scores by race/ ethnic group, in 1976 (when SAT race/ ethnic data began to be reported separately), 1989, and 1994. Blacks increased their mean SAT-Verbal score by 20 points. Hispanic SAT-Verbal scores increased for both Mexican-Americans (+1), and Puerto Ricans (+3). Asian-Americans increased SAT-Verbal scores by 2 points, while Native-Americans' scores went up 8 points. During the same 19 years, mean SAT-V scores decreased 8 points for whites.

Table 2 also displays the differences in mean scores in mathematics for race/ethnic groups in 1976, 1989, and 1994. White mean scores rose 8 points. Over the same period, Blacks gained 34 points, Asian-Americans 17 points, Native-Americans 21 points, Mexican -Americans 17 points, and Puerto Ricans 10 points. Overall, SAT score trends since 1976 show improving Mathematics scores, particularly for minorities, with relatively stable Verbal scores for majority SAT takers and substantial Verbal score increases for Black and Mexican-American examinees.

National Assessment of Educational Progress (NAEP) Score Trends. Unlike the SAT, the NAEP measures "achievement," or skills and levels of knowledge possessed by students at various ages (9, 13, and 17 years old). Called "the nation's report card," NAEP scores are not used for individual assessment, but rather are aggregated by schools and higher levels for educational policy use. Since 1969, the

NAEP has collected data on America's school children in various curriculum areas such as mathematics,

Table 3. Mean Scholastic Assessment Test (SAT) Verbal and Mathematics Scores by Race/Ethnic Group: 1976, 1989, and 1994. science, reading. NAEP scores are scaled from 0 to 500, with various levels identifying successful performance of typical tasks (e.g., 150, 250, 325, and 375). Table 4 displays NAEP Mathematics, Science, and Reading scores from 1969 through 1992.

Race/Ethnic Group	Year					
	1976		1989		1994	
	Verbal	Math	Verbal	Math	Verbal	Math
Native-American	381	420	384	428	389	441
Asian-American	414	518	409	525	416	535
Black	332	354	351	386	352	388
Mexican-American	371	410	381	430	372	427
Puerto Rican	364	401	360	406	367	411
White	451	493	446	493	443	501

Table 4. Mean National Assessment of Educational Progress (NAEP) Mathematics, Science, and Reading Scores: 1 973 - 1992 (Tested Years).

Test Year	Math	Science	Reading	Test Year	Math	Science	Reading
1973	304.4	304.8	285.4	1982	298.5		285.8
1977		295.8	286.1	1986	302.0	283.3	288.8
1978	300.4			1990	305.0	288.5	290.1
1981		289.8		1992		290.0	290.0

Military Selection Testing. The U. S. military has used a composite of ASVAB test scores (Armed Forces Qualification Test [AFQT]) for enlisted selection since 1950. AFQT is a measure of general cognitive ability predictive of performance across a wide set of military training courses. It has included tests of verbal and quantitative ability, plus on occasion, measures of tool knowledge and space perception. Regardless of test content, it has been scaled on a 1 through 99 percentile scale calibrated back to the original AFQT metric. Figure 4 displays the AFQT score trend for military recruits from 1952 to 1994. The dip in AFQT scores between 1976 and 1981 reflects a miscalibration of ASVAB scores which occurred between fiscal years 1976 through 1980. Due to the error, scores of relatively low-scoring applicants were over-estimated, thus over-estimating the ability levels of recruits. The overall pattern of AFQT scores of military recruits through the four decades displayed in Figure 1 consistently reflects (excluding the miscalibration period) increased general ability. In general, current military recruits are among the most able, in terms of AFQT scores, in our history. This is fortunate, because higher technology jobs, a diminishing force requiring a broader set of skills, and greater competition from industry for military-age youth are increasing in relative impact. These personnel ability needs and changes in the demographic composition of American youth have important implications for future military selection and classification policy.

Future Trends

Major demographic changes in the American youth population are occurring. Table 5 shows the 1990 composition of the U.S. population by race/ethnic group, and projections for 2005 and 2015, based upon the 1990 census. The data clearly show that between 1990 and 2015, the proportion of minorities in the American population will grow by about one-third, from 28.2 percent in 1990 to 37.2 percent 25 years later. It should be noted that these percentages are for the entire U. S. population, not just military-eligible youth. In the 18-25 age cohort, minorities will represent 44.5 percent, an increase from 32.9 percent in 1990 (Kageff & Laurence, 1994). Given the generally lower AFQT scores of minorities and projected increase in the number of English-as-a-second-language recruits, military selection, classification, and training will be affected. It is essential that military policy makers consider the long-term implictions of the changes in the pool from which military applicants will be drawn.

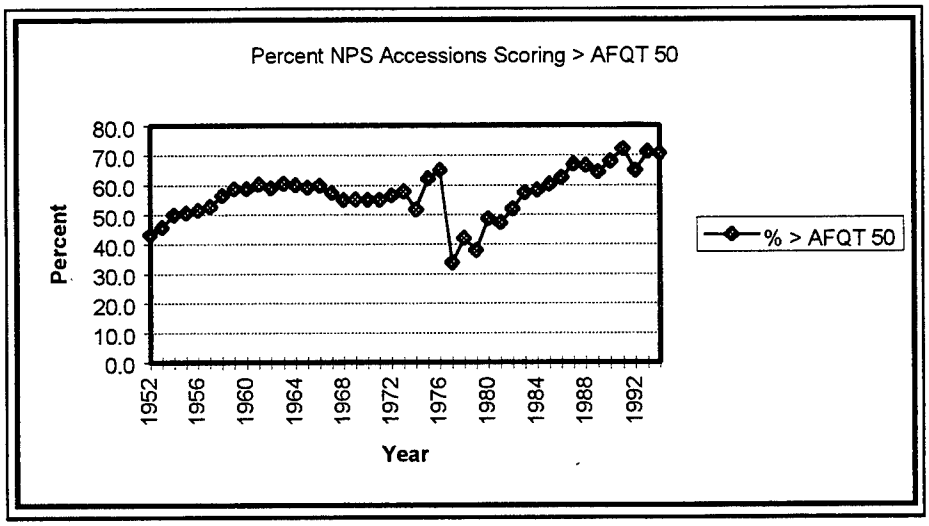


Figure 1. Percentage of Military Recruits Scoring AFQT 50 or Above: 1952 to 1994.

Table 5 . Composition of the U. S. Population by Race/Ethnic Group: 1990, 2005, and 2015

Race/Ethnic Group	1990		2005		2015	
	Number	Percent	Number	Percent	Number	Percent
White	180,250	71.8	187,436	66.4	190,702	62.8
Black	29,348	11.7	36,557	12.9	41,943	13.8
Hispanic-American	22,522	9.0	34,792	12.3	44,450	14.7
Asian-American	7,329	2.9	10,627	3.8	13,284	4.4
Native American	1,962	0.8	2,256	0.8	2,436	0.8
Other	9,451	3.8	10,665	3.8	10,629	3.5
Total	250,862	100.0%	282,333	100.0%	303,444	100.0%

References

Bureau of Labor Statistics (1989). Handbook of labor statistics. Bulletin 2340. Washington, DC: Superintendent of Documents, U.S. Government Printing Office.
Camara, W. J. (1995). Personal communications.

Department of Defense (1996). Population representation in the U. S. military: Fiscal year 1994. Washington, DC: Office of the Assistant Secretary of Defense (Force Management Policy).

Department of Defense (1982). Profile of American youth: 1980 nationwide administration of the Armed Services Vocational Aptitude Washington, DC: Office of the Assistant Secretary of Defense (Manpower, Reserve Affairs, and Logistics).

Kageff, L.L., & Laurence, J.H. (1994). "Test score trends and the recruit quality queue." in Eitelberg, M.J. & Mehay, S.L. (1994). Marching toward the 21st century: Military manpower and recruiting. Westport, CT: Greenwood Press.

Quality of Life in the Navy[4]

Gerry L. Wilcove, Ph.D.
Navy Personnel Research and Development Center
J. Philip Craiger, Ph.D.
University of Nebraska at Omaha
Joyce Shettel Dutcher, Ph.D.
Navy Personnel Research and Development Center

Abstract

This paper presents survey results and structural equation modeling (SEM) results from a quality of life (QOL) questionnaire distributed Navy-wide. Survey results pertained to 11 life domains, overall QOL, conflict or problem areas, and QOL correlations with organizational outcomes. SEM results pertained to 10 models constructed on the basis of demographic subgroups in which overall QOL served as the dependent variable. Parameter estimates were obtained between overall QOL and organizational outcome variables.

Although use of the term "quality of life" (QOL) is pervasive in both the military and civilian sectors, consensus is lacking on the precise definition of the term. Perhaps the clearest definition is the one proposed by Rice (1994, p. 157): "The quality of life is the degree to which the experience of an individual's life satisfies that individual's wants and needs (both physical and psychological)." In the Navy, QOL is often conceived as being a function of eight "pillars": (1) work factors, (2) compensation, (3) personnel policies/practices, (4) medical care, (5) housing, (6) family and individual support services, (7) morale, welfare, and recreational services, and (8) command and personal excellence factors that include leadership, equal opportunity values, voluntary education, fitness, a drug-free work environment, and so forth.

Kerce (1992, 1995) has stated that a variety of reasons exist for studying QOL. She reasons that the all-volunteer Navy must compete with private industry for a decreasing number of eligible recruits. This task is complicated by the fact that prospective recruits are better educated than their predecessors and thus have more alternatives available to them, are more aware of alternatives because of improved access to information, and, with increased alternatives, have higher expectations. Given this climate, the Navy must offer an attractive QOL in order to attract and retain qualified personnel.

In addition, because of decreasing funds, prioritizing becomes crucial. Research can help in at least three ways: (1) it can identify problem areas, (2) determine the relationship of need satisfaction with global QOL, with need satisfaction being examined for specific life domains or QOL pillars, and (3) determine the relationship between global measures of QOL with organizational outcomes, such as intention to remain.

---

[4]The views expressed in this paper are those of the authors, are not official, and do not necessarily reflect the positions of the Navy or the Department of Defense.

This paper presents the results of two sets of analyses. The first set examined survey responses to determine the degree to which QOL was satisfactory to naval personnel. The second set of analyses (Craiger, Weiss, Butler, & Goodman, 1995) used structural equation modeling to identify the models that best fit the survey data for a variety of demographic groups.

Method

Sample

A random sample of 15,000 naval personnel were mailed questionnaires. A total of 7,100 surveys were returned, a response rate of 47%. The composition of the return sample was as follows: officer (26%) and enlisted (74%), shore-based (56%) and afloat-based (44%), male (82%) and female (18%), married (65%) and single (35%), personnel with children (50%) and without children (50%).

Measures

Based in part on the research literature and in part on Navy concepts, the survey examined life domains, areas of possible conflict, global QOL, and organizational outcomes. Life domains included (among others) work, professional development, relationships with children, location (city/town), health care, and pay. Degree of conflict was measured by single items concerned with childcare, medical care, deployments, length of working hours, and so forth. Overall QOL items compared (1) Navy life with civilian life, and (2) current experiences with "what should be" (congruity), as well as simply asking about overall QOL ("Global QOL"). Organizational outcomes included intention to remain, personal readiness, and self-rated performance. Items (especially those concerned with life domains) were formed into scales, where justified, based on Chronbach reliability analyses.

Analyses

To analyze survey responses, SPSS (Statistical Package for the Social Sciences)(Version 6.1) was used to conduct factor analyses, Chronbach reliability analyses, analysis of variance, crosstabulations, and descriptive analyses. Aggregate commands were used to determine the percentage of individuals expressing favorable, unfavorable, and neutral responses across items comprising each scale. Cohen's (1992) concepts of practically significant differences were used to evaluate results. The EQS technique was used for the modeling analyses (Craiger et al., 1995). The "comparative fit index" and the "root mean square residual" were used to evaluate practical significance.

Results

Survey Results

Figure 1 presents total sample results for the 11 life-domain scales that yielded satisfactory reliability results. Domains are ordered from high to low by favorable-opinion percentage. For example, the "relationship with children" domain is presented first, with 70% of the total sample offering favorable opinions, 15% offering unfavorable opinions, and 15% stating they were neutral. A majority of individuals were favorable in seven of the domains, while a minority were favorable in four, with only 29% being favorable on the issue of pay.

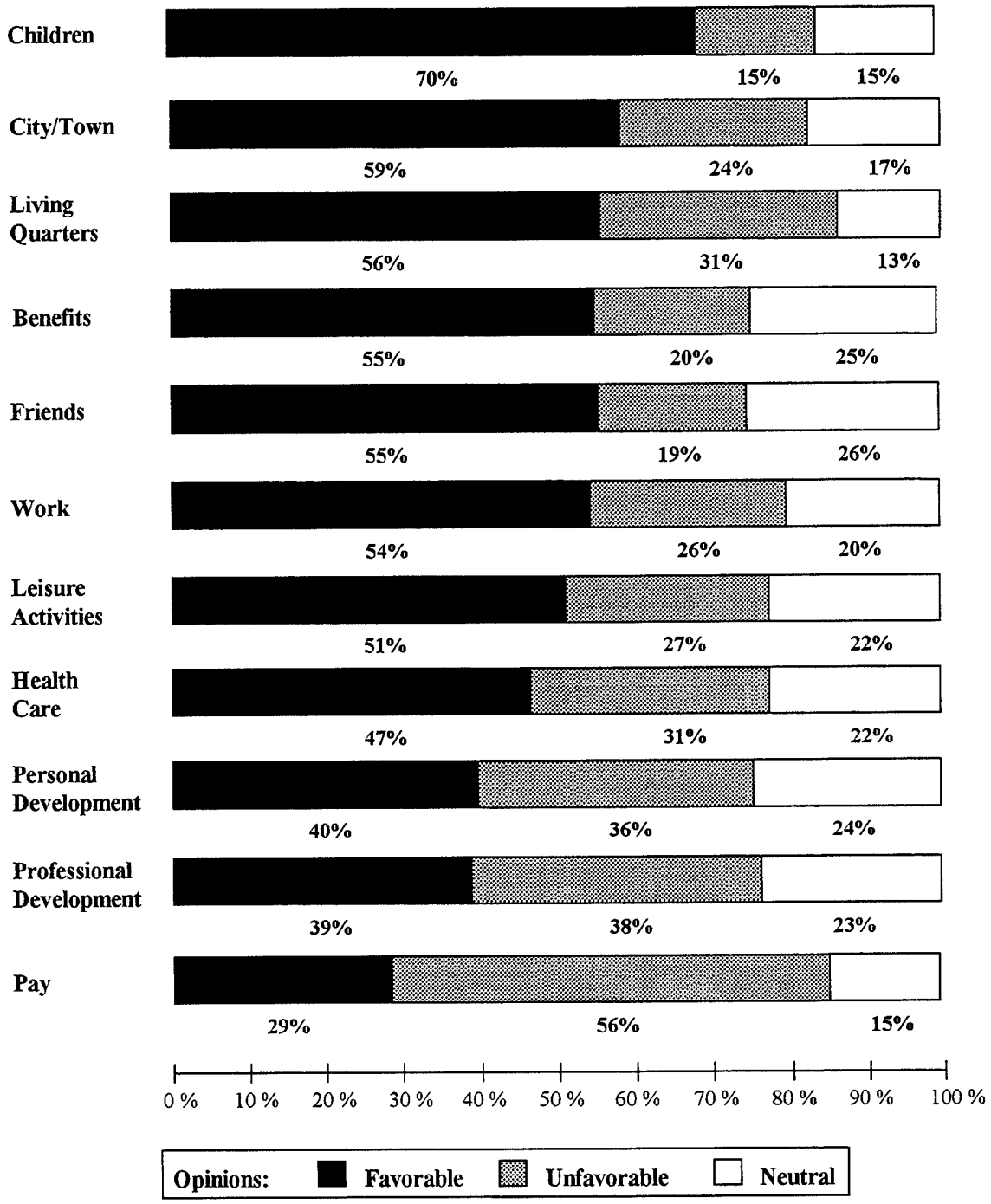


Figure. Subjective quality of life in specific life domains: Total sample results.

Officers were more favorable than enlisted personnel in six of the life domains, the greatest disparity being on the issue of pay, with 54% of officers, but only 21% of enlisted personnel being favorable. Enlisted personnel, in particular, were more favorable when onshore than at sea. For example, 51% of those onshore, but only 26% of those at sea were favorable regarding personal development. Aside from rank, few if any statistically and practically significant differences were found by gender, parental status, and marital status.

Only 18% of the total sample reported that military life was better than civilian life, and only 48% expressed favorable opinions on Global QOL. In an apparent contradiction, 62% felt that Navy life was what it should be or better.

Forty-three percent of personnel reported that their lives were pretty much free of conflicts or problems, 39% indicated that they faced serious problems, and 18% stated that they were experiencing moderate problems. No statistically and practically significant demographic differences were found.

Almost all the correlations between overall measures of QOL and organizational outcomes were below .30. The only exceptions were found for E-2s through E-5s. Results showed that the more favorable such individuals were towards military life compared with civilian life, the more likely they were to want to remain in the Navy. Specifically, the "civilian-military" scale was correlated .41 and .40 with two different items on career motivation.

Modeling

Seventy to 75% of the variance in Global QOL was accounted for by the life domains when models were tested for 10 demographic groups. All of the models were statistically and practically significant.

A number of predictive results were found across demographic groups when examining individual life domains. First, the strongest predictor of Global QOL was work satisfaction, with a weighted (standardized) estimate of .35. The largest (most predictive) estimate was .44 found for men, while the lowest was .24 found for women. The second strongest predictor was satisfaction with leisure activities which produced an estimate of .21. The highest estimate was .30 found for women, and the lowest was .20 found for parents.

Conflict was negatively related to Global QOL, intention to remain in the Navy, and readiness. Favorable perceptions of civilian life were negatively related to Global QOL, intention to remain in the Navy, and readiness. Congruity between actual life experiences and "what should be" was positively related to Global QOL. Global QOL was positively related to intention to remain in the Navy and readiness. Readiness was positively related to intention to remain in the Navy. Conflict was negatively related to both perceptions of civilian life and congruity. Perceptions of civilian life were negatively related to congruity.

Discussion

Total sample results suggest that QOL needs to be improved in the Navy. That is, favorable responses numbered less than 60% in all but one of the 11 life domains. "Personal" domains were rated least satisfactory, the percentage of favorable responses being 51% for activities, 47% for health care, and 40% for personal development.

Additional work is needed (perhaps group discussions) to reconcile the results found for overall measures of QOL. Specifically, why did 62% of the individuals believe that their lives were what they should be or better, while only a minority expressed favorable opinions about Global QOL and that military life was better than civilian life?

Modeling identified the domains that were most correlated with Global QOL. As such, the Navy is in a better position to establish meaningful priorities for allocating funds. However, it should be noted that some domains yielded low parameter estimates, not because their correlations with Global QOL were low, but because they were highly correlated with other, strongly predictive domains.

Conventional correlational analyses yielded .40 coefficients for enlisted personnel between civilian-military comparisons and two measures of career motivation. In contrast, modeling yielded a parameter estimate of .09 with the only career motivation item examined for enlisted personnel. While the two statistics are not directly comparable, the same yardstick can be used to evaluate their practical significance--.40 or greater is large and .10 or lower is very small. Why the difference? Perhaps, the .40 coefficients were inflated due to systematic measurement error. Further, perhaps the parameter estimate was reduced, because inclusion of senior enlisted personnel restricted response variation on the issue of career motivation.

References

Cohen, J. (1992). Statistical power analysis. In Current Directions in Psychological Science, 1, 98-101.

Craiger, J. P, Weiss, R. J, Butler, A, & Goodman, D. (1995). Navy quality of life predictive model project: Results of second administration (TCN Number: 95020). Omaha, NE: University of Nebraska.

Kerce, E. W. (1992). Quality of life: Meaning, measurement, and models (NPRDC TN 92-15). San Diego: Navy Personnel Research and Development Center.

Kerce, E. W. (1995). Quality of life in the U.S. Marine Corps (NPRDC TN TR95-4). San Diego: Navy Personnel Research and Development Center.

Rice, R. W. (1994). Work and the quality of life. In S. Oskamp (Ed.), Applied psychology, annual 5: Applications in organizational settings (pp. 155-177). Beverly Hills: Sage.

The Senior Leader Equal Opportunity Survey: What Do the Bosses Believe?[5]

M. R. Dansby, Ph.D.
Defense Equal Opportunity Management Institute

Abstract

This paper presents the psychometric properties and results of the Senior Leader Equal Opportunity Climate Survey. Over 500 generals, admirals, and Senior Executive Service civilians were surveyed. The instrument demonstrated acceptable reliability. Results indicate senior leaders hold an optimistic view of equal opportunity in the Department of Defense.

In March of 1994, Secretary of Defense Perry issued a memorandum establishing several equal opportunity (EO) initiatives in the Department of Defense (DoD). Among these is the requirement that all newly selected general and flag officers (O7s) and Senior Executive Service (SES) civilians receive a two-day EO training seminar conducted by the Defense Equal Opportunity Management Institute (DEOMI). The author developed the Senior Leader Equal Opportunity Survey (SLEOS) to be used in the seminars to facilitate discussion of EO issues. In addition to a number of new scales, the SLEOS includes several that are comparable to scales in the Military Equal Opportunity Climate Survey (MEOCS; Dansby & Landis, 1991; Landis, Dansby, & Faley, 1993), which has been used across the DoD since 1990 to aid military commanders in identifying and addressing EO and organizational effectiveness concerns.

Between March and November 1995, 20 senior leader seminars were conducted, with the SLEOS administered to all participants. Although only new O7s and SES members are required to attend, a significant number of higher ranking general/flag/SES individuals availed themselves of the seminars. Since only one other senior leader EO survey is presented in the literature (a survey of Navy admirals reported by Gentner, 1986, which included no analysis of the psychometric properties), the SLEOS offers a unique "view from the top" perspective on EO issues within DoD.

Some preliminary analyses of SLEOS results using a smaller database (Hochhaus, 1995; Johnson, 1995; McIntyre, 1995) indicated senior leaders hold a generally optimistic view of the

---

[1]The opinions expressed are those of the author and do not reflect the official views of the DoD or any of its agencies[6] We wish to thank Mr. William Fulton, Clerk of Court of the U.S. Army for providing the data used in this study. This study was completed while the first author was Visiting Professor at the Defense Equal Opportunity Management Institute. The opinions in this paper are those of the authors and do not necessarily represent those of the U.S. Government, the Department of Defense, or their agencies. Request for reprints should be sent to the first author at: Center for Applied Research and Evaluation, University of Mississippi, University, MS, 38677. USA. (e-mail: ijir@vm.cc.olemiss.edu).

DoD's EO climate. Johnson (1995) reported 5-point scale scores ranging from 4.0 to 4.7 (higher is better) for perceptions of fairness, personal preparation, mission relatedness, value of training and assessment, and leadership impact in EO. McIntyre (1995) developed seven scales (alphas ranging from .59 to .82) using only the EO perceptions section of the survey and demonstrated moderate evidence of convergent validity with MEOCS for four of the scales. Hochhaus (1995) conducted a content analysis of responses from almost 250 senior leaders to the open-ended section of SLEOS. He found they identified the following (in order) as the most significant EO issues facing the Services today: opportunities for promotion, retention, etc.; sexual harassment; "reverse" discrimination; general EO issues; recruiting; training; racial discrimination; affirmative action; women in combat or at sea; and downsizing. Each of these was mentioned by at least 10% of the respondents. Over 60% mentioned leadership as the key to an effective EO program, and over 40% mentioned training as being a key element to success.

The following report describes psychometric properties of the SLEOS and summarizes the results from over 500 senior leaders. Based on MEOCS (Dansby & Landis, 1991; Landis, Dansby, & Faley, 1993) results indicating significant differences in EO perceptions by race, gender, and personnel category, the SLEOS data are also examined for such effects.

Method

An initial draft of the SLEOS was constructed, and preliminary field tests using faculty and students at DEOMI resulted in a form with 95 closed-ended and 6 open-ended items. The closed-ended items included 18 demographic items, 25 items measuring general EO perceptions (EOP), 16 items addressing the seriousness of EO issues (EOI), 24 items from MEOCS scales, and 12 leadership scale items using Fiedler's (1967) Least Preferred Coworker (LPC) technique. The closed-ended EO items used 5-point Likert scales with various anchor points. The open-ended items asked for views on significant EO issues in the DoD, strengths and weaknesses of EO programs, elements in an effective EO program, and any other comments relating to EO.

The SLEOS was constructed with several criteria in mind. First, to have comparability with MEOCS (which has a database of over 3200 unit administrations and 400,000 military and civilian respondents throughout DoD and the Coast Guard), several abbreviated scales from MEOCS were included. These scales had been used in experimental versions of MEOCS, and internal analyses had shown them to have acceptable psychometric properties (confirmed by reanalysis during the present study; results are reported in a later section). The following six MEOCS scales (see Dansby & Landis, 1991; Landis, Dansby, & Faley, 1993) were used:

Sexual Harassment & (Sex) Discrimination	Differential Command Behavior toward Minorities
Positive Equal Opportunity Behaviors	Racist/Sexist Behaviors
"Reverse" Discrimination	Overall Equal Opportunity Climate

A second criterion was to tap key issues identified by Gentner (1986) in his survey of admirals. Most of the EOP and EOI items reflect these issues, as do the open-ended items. A third criterion was to include a measure of leadership, in order to examine the relationship between senior leaders' orientation toward leadership and their EO views. The LPC measure was selected because of its brevity, simplicity, and relative lack of demand characteristics.

The survey was administered by mail, approximately three weeks before the seminars began. Each instrument included a cover letter describing the purpose and uses of the survey; a booklet including the privacy act statement, instructions, and survey items; a computer-scored response form for the closed-ended items; and a return envelope. Respondents were advised that the survey is voluntary, but that the overall results would be used as an integral part of their training. They were also assigned a confidential identification code so that individual response profiles could be returned privately to them during the seminar. They were asked to complete and return the response form and comments (if any) to the open-ended items.

Completed surveys were analyzed and reports generated for each seminar. In addition to a summary of results for their class and previous classes, respondents received individual profiles contrasting their views with average scores for the class and previous classes. The results from all classes were included in an overall database, upon which the present analysis is based.

A total of 512 useable survey forms were returned from the 661 seminar participants, yielding a useable response rate of 77%. Demographics for the sample were as follows: 33% Air Force, 17% Army, 32% Navy, 15% other federal civilian; 90% men, 10% women; 93% majority, 7% minority (of which 4.5% were Black and 1.4% Hispanic); 69% military, 30% DoD civilian; of the military members, 37% O7 selects, 38% O7, 25% O8 and above, 70% active duty, 30% National Guard/Reserve duty; of the DoD civilians, 52% SES1, 23% SES2-3, 25% SES4 or higher.

Results

A principal components factor analysis of the 65 EO items resulted in 15 factors with eigenvalues greater than one, accounting for 67% of the variance. The results confirmed the structure for the MEOCS scales and established the structure for nine new scales. Table 1 presents the results of a reliability analysis using Cronbach's alpha coefficient for scales constructed from each factor. The data in Table 1 suggest 12 useable scales (Factors 1-12; psychometric properties of Factors 13-15 are considered unacceptable). Table 2 presents summary statistics for these scales (scale scores computed as average item scores).

Table 1

Factor Loadings

(Rotated Factors; Varimax Rotation)

FACTOR	TITLE	NUMBER OF ITEMS	RANGE OF FACTOR LOADINGS	AVERAGE FACTOR LOADING	STANDARDIZED ALPHA FOR SCALE
Factor 1	EO Issues	14	.64 to .82	.75	.95
Factor 2*	Differential Command Behavior toward Minorities	5	.65 to .81	.74	.89
Factor 3*	Positive EO Behaviors	5	.76 to .86	.82	.90
Factor 4	Success of EO Programs	5	.55 to .83	.72	.85
Factor 5	Helpfulness of EO Programs	5	.52 to .71	.62	.70
Factor 6	EO Link to Leadership and Readiness	6	.42 to .68	.57	.74
Factor 7*	"Reverse" Discrimination	4	.65 to .82	.76	.81
Factor 8*	Racist/Sexist Behaviors	4	.66 to .81	.74	.81
Factor 9*	Sexual Harassment & (Sex) Discrimination	4	.45 to .74	.63	.83
Factor 10	Relative EO Climate in DoD	3	.47 to .80	.68	.68
Factor 11	Concerns about Preferential Treatment for Women & Minorities	2	.81 to .75	.78	.86
Factor 12*	Overall EO Climate	2	(both .72)	.72	.90
Factor 13	Comfort with Personal EO Knowledge	2	.52 to .78	.65	.39
Factor 14	Importance of Commander's Leadership to EO	1	-	.46	-
Factor 15	Need to Handle EO within the Chain of Command	2	.47 to .79	.63	.29

*Abbreviated MEOCS factors

Table 2

Factor Scale Score Statistics (higher score is better)

FACTOR	TITLE	MEAN	STANDARD DEVIATION	N
Factor 1	EO Issues	4.02	.59	510
Factor 2	Differential Command Behavior toward Minorities	4.51	.65	512
Factor 3	Positive EO Behaviors	4.27	.76	512
Factor 4	Success of EO Programs	4.17	.70	511
Factor 5	Helpfulness of EO Programs	4.03	.62	511
Factor 6	EO Link to Leadership and Readiness	4.48	.49	512
Factor 7	"Reverse" Discrimination	4.20	.73	508
Factor 8	Racist/Sexist Behaviors	4.39	.69	511
Factor 9	Sexual Harassment & (Sex) Discrimination	3.87	.87	512
Factor 10	Relative EO Climate in DoD	3.90	.68	511
Factor 11	Concerns about Preferential Treatment for Women & Minorities	4.08	.78	510
Factor 12	Overall EO Climate	4.19	.75	512

A MANOVA was conducted using the 12 factor scales as dependent variables and racial-ethnic category (minority/majority), gender, and personnel status (military/federal civilian) as the independent variables. None of the interactions was significant at the .05 level, nor was there a significant main effect for the military/civilian classification. The racial-ethnic ($p = .000$) and gender ($p = .029$) main effects were significant. A summary of the significant univariate F tests for these main effects is presented in Table 3.

Table 3

Univariate F Tests

(df = 1, 496)

FACTOR	TITLE	Mean Minority (n=34)	Mean Majority (n=463)	p	Mean Women (n=50)	Mean Men (n=447)	p
Factor 2	Differential Command Behavior twd Minorities	3.61	4.51	.000	3.80	4.32	.007
Factor 4	Success of EO Programs	3.25	4.04	.000	3.46	3.83	NS
Factor 5	Helpfulness of EO Programs	4.38	3.96	.019	4.21	4.13	NS
Factor 6	EO Link to Leadership and Readiness	4.73	4.45	.050	4.66	4.51	NS
Factor 7	"Reverse" Discrimination	4.56	4.12	.050	4.42	4.25	NS
Factor 9	Sexual Harassment & (Sex) Discrimination	3.56	3.78	NS	3.40	3.94	.038
Factor 10	Relative EO Climate in DoD	3.19	3.75	.004	3.26	3.68	.030
Factor 12	Overall EO Climate	3.58	4.16	.010	3.79	3.94	NS

Discussion and Conclusions

The present study offers support for use of SLEOS as a tool for assessing senior leaders' views on equal opportunity issues. It also confirms that senior leaders generally have a positive view of the status of EO in the DoD, and that they see a strong link between effective leadership, EO, and readiness. The scale scores that are comparable to MEOCS indicate the generals, admirals, and SES civilians perceive their organizational environments as much more conducive to EO than do service personnel at large. (The only exception is sexual harassment, where the senior leaders and other personnel have similar ratings.) This disparity raises the possibility that senior leaders' positive perceptions may lead them to consider EO a "solved problem" and reject the need for aggressive action to enhance the EO climate.

It is interesting that senior leaders who are racial-ethnic minority members are much more likely to recognize there is still room for improvement in EO. Similarly, senior women are less sanguine than senior men concerning EO issues. These highly successful minorities and women perceive the DoD's EO programs as very necessary, yet not fully successful in bringing about EO results. (Though their views of EO in DoD are generally positive, they are much less so than their majority and male counterparts.) This finding argues for a continuing need to have minorities and women represented in greater numbers at the senior levels of DoD, to encourage greater leadership awareness of EO issues and the need for continued emphasis on EO efforts.

References

Dansby, M. R., & Landis, D. (1991). Measuring equal opportunity in the military environment. International Journal of Intercultural Relations, 15, 389-405.

Fiedler, F. (1967). A theory of leadership effectiveness. New York: McGraw-Hill.

Gentner, F. C. (1986). Navy Flag Officer Pretraining Assessment Survey. Proceedings: Psychology in the Department of Defense Tenth Symposium, Colorado Springs, CO: USAF Academy Department of Behavioral Sciences and Leadership, 576-580.

Hochhaus, L. (1995). A content analysis of written comments to the Senior Leader Equal Opportunity Survey (SLEOS) (DEOMI Research Series Pamphlet 95-7). Patrick AFB, FL: Defense Equal Opportunity Management Institute.

Johnson, J. L. (1995). A preliminary investigation into DEOMI training effectiveness (DEOMI Research Series Pamphlet 95-8). Patrick AFB, FL: Defense Equal Opportunity Management Institute.

Landis, D., Dansby, M. R., & Faley, R. H. (1993). The Military Equal Opportunity Climate Survey: An example of surveying in organizations. In P. Rosenfeld, J. E. Edwards, & M. D. Thomas (Eds.), Improving Organizational Surveys: New Directions, Methods, and Applications. Newbury Park, CA: Sage Publications.

McIntyre, R. M. (1995). Examination of the psychometric properties of the Senior Leader Equal Opportunity Survey: Equal opportunity perceptions (DEOMI Research Series Pamphlet 95-6). Patrick AFB, FL: Defense Equal Opportunity Management Institute.

The Effects of Race on Procedural Justice: The Case of the Uniform Code of Military Justice[6]

Dan Landis, Michael Hoyle, and Mickey R. Dansby
Defense Equal Opportunity Management Institute

Abstract

This research examined potential racial bias in time-related variables inherent in the administration of courts-martial under the Uniform Code of Military Justice (UCMJ). The sample consisted of a database of all charges in the US Army of aggravated assault, drug-related, and sex-related crimes found worthy of prosecution as courts-martial under the UCMJ between 1987 and 1995. Results indicated that blacks were older than whites on non-sex-related crimes, have been in the service longer, and spent longer going from initial charges to final disposition. The relationship was reversed for sex-related crimes. These results were interpreted in terms of an interaction between the level of potential public interest in a crime and the race of the accused, with blacks receiving accelerated treatment in crimes involving sex and less attention in the case of other crimes.

Much theorizing about fairness in the justice system has distinguished between two aspects: distributive and procedural justice (Deutch, 1985; Lind & Tyler, 1988). The first is concerned with the equitable distribution of outcomes (e.g., length of incarceration). The second focuses on equitability in the process as an accused moves through the system to a disposition. The processes may be independent. That is, accused individuals may be treated inequitably [e.g., receive less competent counsel, be the recipient of more intense scrutiny by investigators (Norris, Fielding, Kempe, & Fielding, 1992), receive higher bail in the civilian system, etc.] yet receive the same sentence as a person not so treated. (Blumstein, 1982). Conversely, people can be treated essentially the same while in the system, yet have quite different outcomes. It is the latter aspect that many have focused on to show that the civilian justice system is or is not racially biased. However, the results of such analyses have been mixed with most recent commentators coming to the belief that while there may exist disparities in certain localities and for certain offenses, the civilian justice system is not institutionally biased in the distribution of outcomes (Landis & Dansby, 1994; Tonry, 1995).

Receiving relatively little attention is the potential racial bias that occurs while the accused is being processed through the system. These effects are likely to be subtle, reflecting the amount of attention functionaries in the system give the individual case. Unless the case has warranted a high level of visibility, it may be left to the vagaries of the system to determine the rate of progress through the justice maze. There are a number of methodological reasons for the dearth of research in this area, which cannot be discussed here. Suffice it to say that most studies of procedural justice have not used good dependent measures in their analyses. One direct measure is the time that an accused spends in the system. Since time may be related to actual confinement (as in pre-trial) or to a sense of a lack of closure, one can make the link to cognitive and affective

states directly. And, if there is a consistent racial disparity that can be explained by a reasonable theory, then one can describe an institutional practice that is real and has consequences.

Both the prosecution and defense have many ways of lengthening or shortening this process. In any case, even under the most restrictive reading, there is still sufficient leeway for disparities to exist. Hence, a full understanding of possible racial disparities in the justice process requires such an analysis and is the focus of the present paper. The venue for the present study is the court-martial system in the U.S. Army.

Two studies (Connelly, 1993; Robinson, 1993) have suggested that blacks disproportionately refuse plea bargains as compared to whites. One purpose of this study is to replicate those findings using a larger sample and secondly to determine if the presence of a plea bargain has an impact on the procedural justice aspects of the case.

The possibility of racial disparities in the administration of the UCMJ has been the subject of some research and discussion over the past few years (see Landis & Dansby, 1994, for a summary of these studies). All of the studies have focused on the effect of race on sentence length (i.e., distributive justice) and have generally been unable to indicate any clear racial bias. A possible confounding factor is that, with the exception of the Connelly and Robinson studies, the researches aggregated over all offenses. If there are racial differences in offense profiles as Tonry (1995) suggests, then such aggregation is unwarranted. The sample used in the present study permits an analysis by offense type and thus is an advance over the previous work.

A secondary purpose of this paper is to examine the role of military tenure on likelihood of involvement with the UCMJ. Knouse (1992), based on a small sample of people incarcerated at the Ft. Leavenworth Disciplinary Barracks, suggested that blacks tend to become involved with the discipline system at an earlier age than whites, a possibility echoed by Edwards & Newell (1994), using a sample of records from the Navy. Due to the small sample size, Knouse aggregated over all offenses. The present sample, which consists a database of charges across the entire Army, allows a more precise test of Knouse's suggestion along two dimensions: Time in Service and Age, as well as exploring the role of offense type.

Method and Procedure

Sample: The sample consists of 5989 courts-martial cases obtained from the Office of the Clerk of Court, United States Army Judiciary. 3509 cases involved white personnel while 2480 were charges levied against blacks. The data set covered all reported charges of Aggravated Assault, Drug Crimes, and Sex Crimes levied on soldiers who entered the service between 1 July 1987 and 31 May 1995.

These data indicate that blacks charged with assault, sex-related crimes, and some drug crimes exceed their proportion in the enlisted Army population (about 30%) by a maximum of 250% and are slightly underrepresented in the three marijuana-related offenses. These figures are somewhat at variance with those occurring in the civilian population(Blumstein, 1982; Tonry,

1995). Comparing the Army and the Blumstein data, it is clear that the military overrepresentation rate is no more than half that of the civilian sector.

Analyses of data: Three variables were formed: length of time in the service (TIS), calculated by subtracting the date of service entry from the date charges were preferred; length of time in the criminal justice system (TCD), calculated by subtracting the date of charges being preferred from the date of hearing conclusion; and, time between charges filed and hearing (TCH). The variables were analyzed using survival analysis and examined separately by offense with race as the independent variable. Censoring was not necessary since only charges that had been adjudicated were included in the analysis. Significance of the difference in survival functions between blacks and whites was assessed by Chi-Square with 1 degree of freedom. The impact of plea bargaining on the length of time in the system was tested by a two-way analysis of variance with race and type of plea bargain as independent variables and TCD as the dependent measure.

Results

Effect of Time in Service: Of the 10 offenses analyzed, 9 produced significant chi-squares for race (Table 1).

Table 1. Means of Time Variables (in Days) by Race and UCMJ Offense (Mean Pairs Bracketed are Significantly Different)

		Offenses										
Variable		nl*	nm	vx	xq	xs	dp	w	yk	lp	rt	qe
TIS***	Race											
	White	922.67	841.97	814.75	773.28	780.11	792.99	761.69	785.81	1074.41	1105.92	1179.47
	Black	1170.97	874.89	913.17	1003.36	959.52	1076.02	931.01	964.16	766.26	563.13	688.39
TCH												
	White	42.19	41.01	34.34	35.88	37.96	43.75	37.55	41.74	52.48	56.06	57.49
	Black	48.51	43.89	36.2	42.08	44.31	48.72	47.01	46.59	51.21	63.64	42.39
TCD												
	White	83.02	75.31	62.76	62.75	66.65	69.4	65.59	68.73	106.7	102.99	112.04
	Black	87.86	81.87	64.75	79.83	71.5	83.58	77.51	84.25	94.9	113.47	79.95

* nl=Aggravated Assault with a Firearm; nm=Aggravated Assault without a Firearm; vx=Wrongful Use of Amphetamine; xq=Wrongful Possession of Marijuana, less than 30 gr.; xs=Wrongful Possession of Marijuana, greater than 30 gr.; dp=Possession of Amphetamine with Intent to Distribute; w=Wrongful Use of Marijuana; yk=Wrongful Distribution of Amphetamine; lp=rape; rt=sodomy; qe=indecent assault.

*** TIS=Time in Service; TCH=Time Between Charges Filed and Hearing; TCD=Time between Charges Filed and Disposition

The offenses fall into two categories: those in which blacks are significantly more senior in service than whites and those where the relationship is reversed. The former consists of the non-sex crimes (e.g., assault and drugs) and the latter involves sexual activities (rape, indecent assault, and sodomy). The differences between these two categories is consistent and striking.

Effect of Time in the Criminal Justice System (TCD): Six of the ten offenses produced significant chi-squares for race (Table 1): Wrongful Possession of Marijuana (less than 30 gr.); Aggravated Assault without a Firearm; Possession of Amphetamines with intent to Distribute; Wrongful use of Marijuana; Rape; and Indecent Assault. In the case of the first four (the non-sex crimes), blacks spent significantly more time in the system than whites; the converse was true for the two sex crimes.

Effect of Plea Bargaining: The two-way analysis of variance (race by plea bargain) produced significant interaction effects in only two of the six offenses (Wrongful use of Marijuana, less than 30 gr. [$F=3.73$, $p<.05$] and Possession of Amphetamines with intent to distribute [$F=7.02$, $p<.001$]). An inspection of these data suggests that the overriding effect is racial, rather than the presence of a plea bargain, although it does appear that accepting a plea bargain with conditions attached results in a longer time in the system for blacks when compared with whites. However, the fact that this effect only occurs in two of eleven charges weakens the conclusion of a consistent racial effect.

Effect of Age: Survival analysis using age as the dependent and race as the independent variable gave results paralleling those from the TIS analysis: on non-sex offenses blacks are significantly older than whites; the reverse is true for the three sex crimes. All effects were significant.

Discussion

This study examined the impact of race on time through the courts-martial system in the U.S. Army. In contrast to studies that have found no racial disparity in terms of sentence length, we found large and significant differences in how long it takes to traverse the system. We also found significant racial differences in both age and military tenure of offenders. We would argue that the means by which adjudication comes about are at least as important as the end result. For the accused who is faced with the task of defending him/ herself, time may be either a friend or a foe. A longer time may provide more opportunities to prepare a persuasive case at trial. It may, conversely, provide pressure on the defendant to accept a less-than-optimum decision in order to obtain closure. We would argue that, for most minority defendants (who may be less likely to have the resources to hire expensive civilian counsel), time is not a friend.

The age of the offender depends very much on the type of offense. Part of the reason for the failure to replicate previous studies may lie in the different databases used. The Knouse study used a small sample of serious offenders serving time at Ft. Leavenworth. In contrast, the Edwards and Newell research concentrated on discharges for misconduct, which includes much more than felony level offenses. The present study used a much more extensive data set--all serious offenders over a fairly long period of time. Hence, our data set is more representative of soldiers who find themselves in trouble than the previous sets and this may explain the differential results.

The differences between sex- and non-sex-related crimes may be explained by stereotypes that prosecutors have about blacks and sexuality. We would suggest three factors: 1) sex related crimes are repugnant to victims, prosecutors, and defense counsel alike, and there may be a reluctance to drag such proceedings out; 2) defense counsels may be reluctant to take on these cases, leaving them to lawyers with less experience, and 3) the stereotype, either explicit or implicit, that blacks are less able to control their sexual impulses leading to a judgment such persons are more than likely guilty of the charges. These hypotheses and those involving non-sex crimes need to be verified by further research.

References

Blumstein, A. (1982) On the racial disproportionality of the United States prison populations. Journal of criminal law and criminology, 73, 1259-1281.

Connelly, J. (1993). Equitability of treatment in the Army judicial proceedings (ETAJUP). (Report SR SR-93-14) Bethesda, MS: U.S. Army Concepts Analysis Agency.

Deutch, M. (1985). Distributive justice: A social psychological perspective. New Haven, CT: Yale University Press.

Edwards, J. E., & Newell, C.E. (1994). Navy pattern of misconduct discharges: A study of potential racial effects. San Diego, CA: Navy Personnel Research and Development Center.

Knouse, S. B. (1993). Differences between black and white military offenders: A study of socioeconomic, familial, personality, and military characteristics of inmates at the United States Disciplinary Barracks at Fort Leavenworth. (DEOMI Research Series Pamphlet 93-2). Patrick AFB, FL: Defense Equal Opportunity Management Institute.

Landis, D., & Dansby, M. R. (1994). Race and the military justice system: Design for a program of action research. (DEOMI Research Pamphlet 94-3). Patrick AFB, FL: Defense Equal Opportunity Management Institute.

Lind, E. A., & Tyler, T. R. (1988). The social psychology of procedural justice. New York: Plenum.

Norris, C., Fielding, N., Kempe, C., & Fielding, J. (1992). Black and blue: An analysis of the influence of race on being stopped by the police. British journal of sociology, 43, 207-224.

Robinson, A. C.(1993). Blacks and the military justice system. Unpublished paper. Washington, DC: HDQR, Department of the Army

Tonry, M. (1995). Malign neglect: Race, crime and punishment in America. New York: Oxford University Press.

The Relationship Between Racism/Sexism and Group Cohesiveness and Performance[1]

Robert E.Niebuhr
Auburn University
Stephen B. Knouse
University of Southwestern Louisiana
Mickey R. Dansby
Defense Equal Opportunity Management Institute
Katherine E. Niebuhr
University of Montevallo

Abstract

Survey data from more than 1000 active-duty military personnel indicated significant relationships between perceptions of racism, sexism, cohesion, and performance. The impact of gender and racial biases on these relationships was also examined.

Group cohesiveness is a key concept in the examination of military unit performance (Mael & Alderks, 1993; Oliver, 1988). Social researchers have defined cohesiveness in a variety of ways: "tendency for a group to be united in the pursuit of its goals" (Carron, 1982, p. 124), commitment to the group (Cartwright & Zander, 1968), and, more subjectively, a "we feeling" of emotional climate (Vraa, 1974). However, personal animosities among group members can be debilitating because they generate friction. In a recent meta-analysis of literature, Mullen and Copper (1994) found that commitment to the group task is the critical component in group cohesiveness.

As work groups are becoming more racially, ethnically, and gender diverse, the influence of workforce diversity on group dynamics is complex. On the one hand, diverse (or heterogeneous) groups may require more time and effort to resolve individual differences in perspectives and approaches to problems. Diversity may inhibit cohesiveness because group members can find fewer commonalities upon which to build mutual goals and supportiveness. For example, Terborg, Castore, and DeNinno (1976) found that groups with less similar attitudes among members reported less cohesiveness than did groups whose members exhibited similar attitudes. Conversely, these differences actually may produce more creative decisions (Thornburg, 1991) and allow the group to deal more effectively with complex problems that require critical analysis and innovative solutions (McCleod, Lobel, & Cox, 1992).

Two factors, racism and sexism, can produce discriminatory climates in work groups and have been shown to be intercorrelated and related to such variables as lower cognitive sophistication and anti-egalitarianism (Sidanius, 1993). Moreover, these two factors cause groups to contrast themselves sharply with a perceived outgroup (e.g., minorities or females) (Henley & Pincus, 1978).

[1]The views presented in this paper are those of the authors and do not represent the official positions of the Department of Defense or any of its agencies.

Several meta-analyses have explored the relationship between cohesiveness and performance. Oliver (1988) found a mean r of .32 with 14 military and civilian field studies, while Evans and Dion (1991) reported a reliability corrected mean r of .42 for 16 field and experimental studies. Recently, Mullen & Copper (1994) reviewed 49 studies, computing a mean r of .25 for these studies. Among their findings was evidence for the directionality of cohesiveness and performance. In addition, their meta-analysis demonstrated that certain factors influenced the cohesiveness-performance relations, such as group size, real groups, and task commitment. The most recent meta-analysis by Gully, Devine, and Whitney (1995) found similar overall results (corrected r = .32 for 46 studies) but also examined differences due to level of analysis (group versus individual) and task interdependence.

The purpose of the present study is to explore the relationships between discriminatory climates of work groups (i.e., the acceptance or non-acceptance of diversity), cohesiveness, and performance in naturally occurring work groups. While the relationship between discriminatory climates and group outcomes has not been specifically studied, theoretical models on attitude dissimilarity may support a negative relationship between discrimination (as a negative attitude) and group cohesiveness and performance (Terborg, et al., 1976).

Method

The data for this study included responses from a sample of 1128 subjects from an active-duty military unit located in the U.S. The instruments for each of the samples contained self-report Likert-type items with five response categories. Anonymity of responses was guaranteed. The instruments used in the study are described as follows:

Discriminatory Climates. The measures of gender discrimination (sexism) and racial discrimination (racism) were obtained from the Military Equal Opportunity Climate Survey (MEOCS) (Landis, Dansby, & Faley, 1993). These scales consist of six behavioral incident items each, rated by the respondents on the probability of the behavior occurring in their unit. Group Cohesiveness. The respondents completed a four-item peer cohesion instrument developed by Siebold & Lindsay (1994), the scale focusing on the "attraction to the group" and "commitment to the group task" criteria emphasized by Mudrack (1989) in a review of the cohesion measurement literature. Factor analysis confirmed that the factor structure of the instrument was unidimensional.

Group Performance. The three-item group performance scale evaluated perceived quality and quantity of group output.

Table 1 presents the descriptive statistics for the data. Factor analyses confirmed the factor structures for the discrimination scales. Table 2 provides the reliabilities of all measures (Cronbach's alpha coefficients on the diagonal) and the intercorrelations among the study variables. The reliabilities for sexism and racism are consistent with those obtained in the development of the original MEOCS instrument (Landis, Dansby, & Faley, 1993).

Table 1

Descriptive Statistics

Variable	Mean	Standard Deviation
Sexism	3.94	0.99
Racism	3.26	1.03
Cohesion	3.40	1.05
Group Performance	2.32	1.04

Results

As shown in Table 2, the correlations among the study variables were significant, thus supporting the hypothesis that discriminatory climates would be negatively related to group functioning (i.e., cohesiveness and performance). Partial correlations among racism, sexism, cohesion, and performance, controlling for overall job satisfaction, were also significant.

Table 2

Correlations Among the Study Variables

	Variable			
Variable	Sexism	Racism	Cohesion	Perf.
Sexism	(0.88)			
Racism	0.50**	(0.85)		
Cohesion	-0.23**	-0.27**	(0.90)	
Group Performance	-0.16**	-0.21**	0.51**	(0.78)

The correlation between cohesiveness and performance (r = .51) is consistent with the mean r's found in the recent meta-analyses of cohesion and group performance studies (Gully, Devine, & Whitney, 1995; Mullen & Copper, 1994, Evans & Dion, 1991; Oliver, 1988).

Discussion

The analysis of the data supports previous findings regarding the relationship between group cohesiveness and performance and, in addition, supports the hypothesized relationship between discriminatory climates and group cohesiveness.

While a number of antecedent factors to group cohesiveness (Lott & Lott, 1965) and to racism and sexism (Sidanius, 1993) have been examined, a further analysis of the data suggests that gender and race of group members may also be important. Analyses of variance for the influence of respondents' race and gender on cohesion and performance found a significant effect of race on

cohesiveness while gender did not have a significant influence. A second set of analyses of variance examined gender and race difference as factors affecting the two discriminatory climates. The non-white group perceived greater racism than did the white group. Likewise, females perceived greater sexism in the environment that did males. It may be that those in a position of less power may be more sensitive to discrimination of any type (Niebuhr & Oswald, 1991). The analyses did indicate that females perceived greater racism climates than did males, and non-whites perceived greater sexism than did whites.

These two post-study analyses support the antecedent variables of group demographics influencing group outcomes. The data only allowed for category comparisons (race and gender across work groups) rather than comparisons of race and gender within groups. While the sexual harassment literature has extensively examined the question of gender mix (Gutek, Cohen, & Konrad, 1990; Gutek & Morasch, 1982; Niebuhr & Boyles, 1991), there has been little research concerning gender mix in the cohesiveness area. Siebold and Lindsay (1994) did examine the influence of group racial mix on perceptions of group cohesiveness and found no effects. It could be argued, however, the Army platoons (their basic level of analysis) are too large for examining actual work group dynamics. Future research should address the race/gender demographics of work units and how they relate to discriminatory behaviors, group cohesion, and performance.

In the present study, the survey data provided an interesting factor which might also be considered in creating a positive environment. The survey asked if the respondent had a close friend of another race. An analysis of this difference indicated a significantly lower perception of racism for those having a friend of another race (versus those that did not have such a friend). Consequently, multi-racial friendships both on and off the job may be a primary means of understanding and hence dealing with racism of the job.

The bi-directionality of the cohesion - performance relationship recently posited offers some possibilities for building cohesion in diverse work groups. For example, the performance to cohesiveness directionality indicated in the Mullen and Copper (1994) meta-analyses would support the idea that successful group performance may produce stronger interpersonal attraction and group pride, which in turn may lead to stronger cohesion. Conversely, early and persistent failures in group performance may lead to blame-placing on certain members with divergent views (e.g., minorities and females) and thus increase perceived racism and sexism. This would imply that early successes in group endeavors would be important for cohesion formation. Team building for diverse work groups should emphasize group work on short-duration tasks carrying a high probability of success early in the development of the group. As cohesion develops, more difficult tasks can then be attempted where the diverse talents of the group member mix can provide a greater pay off.

Future studies should examine diverse cultural work environments and focus upon more objective measures of group performance. In addition, longitudinal studies are needed to refine the causal relationship between cohesion and performance. Given the changing demographics of our society, other discriminatory climates, such as age and disability and their influence on group processes should also be explored. Organizational adaptation to theses changing demographics requires the creation of organizational climates that are conducive to the acceptance of individuals who are "different" from the traditional employee.

References

Carron, A. V. (1982). Cohesiveness in sport groups: Interpretations and considerations. Journal of Sport Psychology, 4, 123-138.

Cartwright, D., & Zander, A. (1968). Group dynamics: Research and theory (3rd Ed.). New York: Harper and Row.

Evans, C. R., & Dion, K. L. (1991). Group cohesion and performance: A meta-analysis. Small Group Research, 22, 175-186.

Gully, S. M., Devine, D. J., & Whitney, D. J. (1995). A meta-analysis of cohesion and performance: Effects of level of analysis and task interdependence. Small Group Research, 26, 497-520.

Gutek, B. A., Cohen, A. G., & Konrad, A. M. (1990). Predicting social-sexual behavior at work: A contact hypothesis. Academy of Management Journal, 33, 560-577.

Gutek, B. A., & Morasch, B. (1982). Sex-ratios, sex-role spillover, and sexual harassment at work. Journal of Social Issues, 38, 55-74.

Henley, N. M., & Pincus, F. (1978). Interrelationship of sexist, racist, and anti-homosexual attitudes. Psychological Reports, 42, 83-90.

Landis, D., Dansby, M. R., & Faley, R. H. (1993). The Military Equal Opportunity Climate Survey: An example of surveying in organizations. In R. Rosenfeld, J. E. Edwards, & M. D. Thomas (Eds.), Improving organization surveys (pp. 210-239). Newbury Park, CA: Sage.

Lott, A. J., & Lott, B. E. (1965). Group cohesiveness as interpersonal attraction: A review of relationships with antecedents and consequent variables. Psychological Bulletin, 64, 259-309.

Mael, F. A., & Alderks, C. E. (1993). Leadership team cohesion and subordinate work unit morale and performance. Military Psychology, 5, 141-158.

McCleod, P., Lobel, S., & Cox, T. (1992). Ethnic diversity and creativity in small groups. Proceedings of the Academy of Management

Mudrack, P. E. (1989). Group cohesiveness and productivity: A closer look. Human Relations, 42, 771-785.

Mullen, B., & Copper, C. (1994). The relation between group cohesiveness and performance: An integration. Psychological Bulletin, 115, 210-227.

Niebuhr, R. E., & Boyles, W. (1991). Sexual harassment of military personnel. International Journal of Intercultural Relations, 15, 445-457.

Niebuhr, R. E., & Oswald, S. L. (1991). The relationship between workgroup composition and sexual harassment: An empirical study. Paper presented at the annual meeting of the Academy of Management, San Francisco.

Oliver, L. W. (1988). The relationship of group cohesion to group performance: A research integration attempt. TR 807. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Sidanius, J. (1993). The interface between racism and sexism. Journal of Psychology, 127, 311-332.

Siebold, G. L., & Lindsay, T. J. (1994). The relation between soldier racial/ethnic group and perceived cohesion and motivation. Paper presented at the annual meeting of the American Sociological Association, Los Angeles.

Terborg, J. R., Castore, C., & DeNinno, J. A. (1976). A longitudinal field investigation of the impact of group composition on group performance and cohesion. Journal of Personality and Social Psychology, 34, 782-790.

Thornburg, T. H. (1991). Group size and member diversity influence on creative performance. Journal of Creative Behavior, 25, 324-333.

Vraa, C. W. (1974). Emotional climate as a function of group composition. Small Group Behavior, 5, 105-120.

Harassment in the Canadian Air Force: 1992 and 1995 Survey Results[7]

Lieutenant(N) Brian R. Thompson, M.A.
Canadian Forces Exchange Officer
Air Force Occupational Measurement Squadron

Abstract

The Canadian Forces (CF) has a zero tolerance harassment policy prohibiting any type of harassment in the workplace. This paper reports Air Command data for the years 1992 and 1995 obtained from the Canadian Forces Personal Harassment Questionnaire (CFPHQ) and discusses the effectiveness of the Air Command Harassment Elimination Programme (HELP) implemented in 1992.

The Canadian Forces Administrative Order (CFAO) on harassment was promulgated in 1988 and states in part that no member shall subject any other member or any other person with whom the member works to any type of personal harassment including sexual harassment. The HELP defines personal harassment, sexual harassment and abuse of authority as follows:

Personal harassment means unsolicited behaviour by an individual that is directed at or is offensive to another individual; that is based on personal characteristics including, for example, race, religion, sex, physical characteristics, or mannerisms; and that a reasonable person ought to have known would be unwelcome.

Sexual harassment means unsolicited behaviour that is directed at or offensive to another individual; that a reasonable person ought to have known would be unwelcome, and that has a sexual purpose or is of a sexual nature. It may include, but is not limited to, unwanted sexual advances, unwanted sexual attention, leering, lascivious or lewd remarks and the display of derogatory material. It consists of actions, remarks, gestures - whether they occur only once or many times - which might be expected to cause offence or humiliation and, notwithstanding the intention of the offender, are unsolicited, unwanted and unwelcome.

Abuse of authority means the misuse of authority to undermine, sabotage, or otherwise interfere with the career of another individual including but not limited to, intimidation, threats, blackmail, coercion, or unfairness in the distribution of work assignments, in the provision of training or promotional opportunities, in the completion of performance evaluations, or in the provision of references.

---

[7] Views expressed in this staff note are those of the author and not necessarily those of the Canadian Forces. Appreciation is extended to Major R.A. Boswell from Air Command, Major J.M. Uchiyama and Captain K.M.J. Farley from CFPARU for their contributions to this paper; Major R.J. Hansen for developing the Canadian Forces Personal Harassment Questionnaire; and survey participants and Wing Personnel Selection Officers who coordinated the survey.

The Canadian Forces Personnel Applied Research Unit (CFPARU) developed the Canadian Forces Personal Harassment Questionnaire (CFPHQ)(Hansen, 1991) and administered it to over five thousand service members in October 1992 just prior to implementation of the Air Command HELP (Hansen, 1993). In March 1995, the CFPHQ was readministered to a random sample of Air Command respondents in order to examine the occurrence of harassment in the Air Force subsequent to introducing HELP. The purpose of this paper is to report the findings of the 1995 survey and to compare the 1995 results with the 1992 CFPHQ data (Thompson, 1995).

Method

Table 1 represents the rank/grade and gender of the sample of the Air Command subjects who completed 1,456 questionnaires (9 did not indicate gender) in October 1992 and 918 (23 did not indicate gender) in March 1995. The 1992 response rate in Air Command was 73%, while the 1995 response rate was 78%. The 1992 sample represents 34% of females and five percent of the males in Air Command, whereas the 1995 sample represents 19% of females and three per cent of the males in Air Command. Females were over-sampled in order to ensure adequate representation since females only represent 12% of the Air Command population. The CFPHQ was administered to voluntary participants in a controlled classroom setting. Upon completion of the CFPHQ, subjects sealed their questionnaires in envelopes for transmission to CFPARU.

Table 1
Survey Respondent Composition

	Females				Males			
Rank/Grade	1992	(%)	1995	(%)	1992	(%)	1995	(%)
Pte to MCpl (E-1 to E-4)	462	(67)	307	(71)	404	(53)	272	(59)
Sgt to CWO (E-5 to E-9)	98	(14)	63	(14)	190	(25)	102	(22)
OCdt to Capt (O-1 to O-3)	117	(17)	54	(12)	121	(16)	66	(14)
Maj and Above (O-4<)	8	(1)	8	(2)	45	(6)	20	(4)
Unknown	2	(0)	3	(1)	0	(0)	0	(0)
Total	687	(100)	435	(100)	760	(100)	460	(100)

Note. The unbracketed number represents the actual number of respondents while the bracketed numbers reflect per cent of the male or female respondents for 1992 or 1995.

Results

Air Command members' awareness of CF harassment policy and their participation in harassment training is presented in Table 2. The 1995 results indicate that over 90% of respondents were aware of the harassment policy, which is an increase of over 10% from the 1992 findings for both males and females. This is supported by the finding that in 1995 approximately half of the Air Command subjects had read the Harassment CFAO. The most

significant difference between 1992 and 1995 surveys is that the percentage of respondents who have attended harassment training significantly increased from 22% to over 80%.

Table 2
Comparison of Air Command Members' Awareness of the CF Harassment Policy

	Female %		Male %	
	1992	1995	1992	1995
Aware of the CF Harassment Policy	80	93	79	90
Read the Personal Harassment CFAO	40	59	35	47
Attended Harassment Training Seminar	22	80	22	83

Table 3 presents percentages of Air Command respondents who indicated that they had been harassed while performing CF duties during the previous 12 months. Data for 1992 and 1995 show that all three types of harassment are prevalent for female respondents, while sexual harassment was reported to be minimal by males. However, the figures in Table 3 show a reduction in all types of harassment amongst respondents from 1992 to 1995. In particular, the number of females reporting sexual harassment in Air Command was 10 per cent lower in 1995.

Table 3
Air Command Members Indicating Harassment During the Past 12-Month Period

Type of Harassment	Female %		Male %	
	1992	1995	1992	1995
Sexual Harassment	27	17	3	2
Personal Harassment (excluding sexual)	31	27	18	15
Abuse of Authority	30	29	24	24

Note. Respondents could indicate more than one type of harassment

In 1995, the most frequent types of sexual harassment for female respondents were: unsolicited and offensive sexual teasing, jokes, remarks, or questions; sexual talk or behaviour that created an offensive, hostile or intimidating environment; and unsolicited/offensive sexually suggestive looks, gestures or body language. For males it was: unsolicited and offensive letters, telephone calls, or materials of a sexual nature; unsolicited and offensive pressure for dates; and offensive attempts to participate in sexually-oriented activities.

The most frequently perceived basis of personal harassment for female respondents were: gender; physical characteristics; and mannerisms while for males it was mannerisms; physical characteristics; and national, regional, or ethnic origin.

The most frequently perceived basis of abuse of authority for females were: blackmail (e.g. threat of a low evaluation); unfairness in the provision of promotional opportunities; and unfair evaluation of their job performance. For males it was: unfairness in the provision of promotional opportunities; blackmail; and unfair evaluation of their job performance.

Response to Harassment and Effect of the Action

Females who reported that they had experienced harassment were most likely to respond by: telling their supervisors; asking the person to stop; ignoring the behaviour; and/or avoiding the person. Males who reported that they had experienced harassment were most likely to respond by: ignoring the behaviour; avoiding the person; going along with the behaviour; and/or telling their supervisors. Responses indicate that avoiding the person, telling their supervisor or asking the person to stop had some positive effect on the harassment situation for both sexes. Generally, ignoring the behaviour made no difference on the harassment situation.

Respondents indicating that they had been harassed were asked whether or not they took action against the perpetrator. In 1992, formal action against the harasser was initiated by 23% of both females and males for all forms of harassment while in 1995, only 11% of males initiated formal action against the harasser and 23% of females chose to take formal action. In 1992, the most commonly reported reaction to formal action was that the supervisor did nothing, whereas in 1995, very few subjects reported that their supervisor did nothing. Respondents who had not taken formal action (1995- 89% males and 77% females), were asked the reason for not doing so. Reasons "I thought it would make my work situation unpleasant" and "I did not think anything would be done" were endorsed most frequently by subjects. The finding that many subjects chose not to report harassment was supported by Aggarwal (1992), who found that 52% of the women he surveyed did not report harassment because they believed that nothing would be done.

Of subjects reporting that they had been harassed, approximately 30% of both female and male respondents for 1992 and 1995 surveys had either requested a posting or considered leaving the military as a result of harassment. In 1995, 20% of female and 22% of the male respondents reporting harassment stated that they were absent from work due to a harassment incident.

Discussion

Comparison of 1992 to 1995 survey results reveals that in 1995, a smaller percentage of respondents believed that they had been harassed in some way while performing their CF duties. This decrease in harassment may be attributed to the subjects' increased knowledge as to what constitutes acceptable behaviour. Perhaps this heightened awareness of the existing harassment policy is due to the fact the majority of respondents had received harassment training. Further, a decline was found in the percentage of members who felt that they had been sexually harassed from 27% of females in 1992 to 17% in 1995 and from 3% of males in 1992 to 2% in 1995. The 1995 results show a slight decrease in personal harassment for both females and males from 1992. However, the majority of females who reported personal harassment felt that their gender was some or all of the basis for this harassment. Although all types of harassment are addressed by the HELP, gender training may continue to require the most concerted effort. It is posited that future

harassment education and increased experience in working with women in non-traditional roles will assist to further reduce the incidence of sexually harassing behaviour.

By definition, abuse of authority is exclusively a superior to subordinate act. The 1995 survey found that 29% of female and 24% of male respondents believed they had experienced some form of abuse of authority during the past 12-month period. Appropriate use of authority is a difficult concept since military leaders are trained to give orders and subordinates are trained to obey them. Although military personnel receive training to deal with difficult circumstances, the way in which highly stressful training or operational situations are handled could be construed as abuse of authority by some subordinates. The most prevalent forms of abuse of authority reported by both male and females were unfairness in the evaluation of performance and in the provision of promotional opportunities. Although any deliberate attempt to undermine an individual's career by underrating would clearly be abuse of authority, past research has found that at least 40% of employees in jobs of all types place themselves in the top 10% performance level (Meyer, 1980). In other words, since they see themselves as the best performers in the organization, it is understandable that they may consider any supervisors' rating as unfair, and could subsequently report inappropriate use of authority. Therefore, complaints of abuse of authority require determination as to whether an appropriate level of authority was used or if complaints are simply a reflection of dissatisfaction with legitimate military practices.

Both 1992 and 1995 surveys indicate that the majority of subjects who believed that they had been harassed did not report the harassment. In fact, less than one quarter of the respondents who indicated harassment had used the formal reporting procedures. As most subjects reported being aware of the formal complaint procedure, this finding indicates either a lack of confidence in the HELP complaint mechanism or the fact that members informally defused the incidents. Subjects who chose to take action outside of the formal harassment complaint system and asked or told the person to stop their behaviour found their actions were more effective in 1995 than in 1992. This finding is consistent with the one of the goals of HELP which is to solve the problem at the lowest possible level. Subjects who took formal action to a harassment incident reported little satisfaction with the follow-up action. Although it is not possible to determine why this occurred, it may be because the resolution/outcome was either not what the complainant had wanted, or because it is difficult to feel positive about the situation regardless of the outcome.

A limitation of this research is that it uses self-report data to measure the occurrence of harassment and as a result fails to consider the fact that intense emotions often associated with harassment may color responses of past events. Despite this, the CFPHQ has the potential to provide sound interpretive data concerning Air Command's harassment policy and provide an accurate estimate of the occurrence of harassment. Since implementation of HELP, the incidence of harassment in Air Command has decreased and the knowledge of harassment policies and procedures have increased. Although eliminating harassment is not an easy process , this paper provides evidence to support the continuance of the HELP with specific attention be given to address abuse of authority and gender integration through future leadership and staff training.

References

Aggarwal, A.P. (1992). Sexual harassment in the Workplace. (Second Edition). Markham, Ontario: Buttersworth Canada.

Hansen, R.J. (1993). Personal Harassment in the Canadian Forces: 1992 Survey. (Working Paper 1/93). Willowdale, Ontario: Canadian Forces Personnel Applied Research Unit.

Hansen, R.J. (1991). Development of the Canadian Forces Personal Harassment Questionnaire. (Technical Note 12/91) .Willowdale, Ontario: Canadian Forces Personnel Applied Research Unit.

Meyer, H. H. (1980). Self-appraisal of job performance. Personnel Psychology, 33, 291-296.

Thompson, B.R. (1995). Harassment in Air Command: 1992 Survey. (Staff Note 95/1). Westwin, Manitoba: Air Command Headquarters.

Hierarchical Classification of Training Needs Affecting Flight Crew Performance

Lawrence L. Bailey, Ph.D. and Rogers V. Shaw, M.S.

Abstract

On October 26, 1993 there was a fatal crash of a Federal Aviation Administration (FAA) flight inspection aircraft. During the accident investigation, the National Transportation Safety Board (NTSB) cited ineffective crew resource management (CRM) as one of the causal factors, and recommended CRM training for flight inspection aircrews. As part of the FAA's response to the NTSB's recommendation, a CRM training needs analysis was conducted. Cluster analytic results of the identified training needs suggested three categories affecting crew performance: (1) technical skills; (2) crew coordination skills; and (3) the organization context in which flight inspection crews perform. Implications for CRM awareness training are discussed.

The training of aircrews has changed markedly from the early days of aviation. Initially, the emphasis of air crew training focused on developing the technical skills of the individual crew members. The underlying assumption was that if crew members were technically proficient at their respective jobs, they automatically would be able to operate effectively as a crew During the 1970s, however, evidence from airline accident reports, flight simulator transcripts, and interviews with crew members suggested that the above assumption was inherently flawed. Technical competence by itself did not insure a successful mission. Instead, mission success was dependent on the manner in which technically competent crews coordinated their individual efforts. Training to develop aircrew coordination skills became known as Crew Resource Management or simply CRM (for a CRM historical review see Hartel & Hartel, 1995). More recently, organizational factors (in particular those which determine the consequence of performance, and those that provide flight crews with the resources necessary to perform their jobs) have also been found to be important determinants of aircrew performance (Hackman, 1993). These findings suggest that the performance of aircrews is dependent on at least three factors: (1) technical competencies; (2) crew resource management skills; and (3) the organizational context in which crew's perform.

Issues surrounding the above performance factors were identified in a National Transportation Safety Board (NTSB) report of the October 26, 1993 fatal crash of a Federal Aviation Administration (FAA) flight inspection aircraft, N82 (NTSB, 1994). In reviewing the factors contributing to the October crash, the NTSB issued one urgent action and seven priority action recommendations to the FAA. Included in the latter was the recommendation to institute Crew Resource Management (CRM) training, as outlined in the FAA CRM Advisory 120-51, at each of the Flight Inspection Area Offices (FIAO).

"Flight inspection" refers to the airborne tests conducted to ensure that airway facilities navigational aids are sending accurate signal-in-space guidance, and to ensure that instrument flight procedures are accurate and will safely guide aircraft to their destination. A flight inspection crew consists of a pilot in command (PIC), a second in command (SIC; co-pilot), and an electronics technician (ET). The flight inspection mission differs from other forms of flying

(such as air transport) in that most flight maneuvers are conducted within the terminal area, at low altitudes, and at times running counter to the established air traffic flow pattern. This requires a high degree of coordination with air traffic control and with the aircraft in order to maintain traffic vigilance.

One of the main emphases of CRM training is to develop the resource management skills necessary to ensure that all crew members are operating from a common frame of reference, and that this reference is consistent with what is actually occurring. Specific skills developed in CRM awareness training commonly include: (1) communication skills such as inquiry, advocacy, and assertion; (2) methodologies for identifying problems and making decisions under severe time constraints; (3) self monitoring skills for critiquing decisions and actions of the crew; (4) conflict resolution skills; (5) skills associated with crew leadership, followership, and concern for the task; (6) interpersonal skills necessary for maintaining a professional crew climate; (7) situational awareness and distraction avoidance skills; (8) workload planning and distribution skills; and (8) identifying personal stressors and developing effective stress reduction techniques (FAA, 1995).

Prior to the accident of N82, the FAA's Office of Aviation Systems Standards (AVN) was in the early stages of developing a CRM program for its flight inspection crews. With the advent of the accident, this initiative was elevated in priority and a CRM task force led by the Civil Aeromedical Institute (CAMI) was created to guide the process of developing a CRM course for the flight inspection mission. One of the first steps taken by the CRM task force was to conduct a CRM training needs analysis based on issues addressed during the November 1993 safety meetings conducted at each of the FIAOs. These meetings were mandated by the FAA Administrator in response to the accident of N82. This paper presents the results of a hierarchical classification of the training needs that emerged from the CRM needs analysis.

Method

Participants

Fifty-eight subject matter experts (SMEs) volunteered to participate in the data collection for a training needs analysis. This represented 30% of the flight inspection workforce. Subjects consisted of PICs, SICs, and ETs. To protect the anonymity of the individuals and their respective FIAOs, no demographic data were collected. In addition, all surveys were destroyed following data entry and analysis. These measures were taken to insure that participants would be candid with their responses, and that no punitive action could result from participation in the needs analysis.

Instrument

Subjects were presented with a questionnaire containing 109 issues that were extracted verbatim from written summaries of the November 1993 safety meeting discussions. For each issue, subjects indicated which of 13 performance categories most applied to a given safety issue. These included: (1) crew interpersonal climate; (2) situational awareness; (3) leadership; (4) communications; (5) mission analysis; (6) workload management; (7) decision making; (8)

adaptability; (9) assertiveness; (10) life stress; (11) skill proficiency; (12) organizational factors; and (13) CRM dimension not specified. Multiple performance categories could be assigned to a given safety issue. Definitions for the first 10 performance categories were derived from commonly accepted CRM dimensions (FAA, 1995). These categories represented potential CRM awareness training modules. Categories 11 and 12 were included based on the literature previously reviewed. Category 13 was included for completeness.

Results

To determine the hierarchical structure of the safety issues a frequency matrix was first developed in which the rows contained the 109 safety issues, and the columns contained the 13 performance categories. Cell values represented the frequency with which a given safety issue was matched to a given performance category. The maximum cell value was 58, corresponding to the number of subject matter experts. Next the frequency matrix was converted into a proximity matrix using squared Euclidean distances as a measure of similarity. Clusters were then formed using the Ward's method in SPSS for Windows version 6.0. Figure 1 shows the resulting hierarchical relationship of five interpretable clusters relating to: (1) technical skills; (2) organizational stressors; (3) crew stressors; (4) situational awareness; and (5) planning and decision making.

Figure 1: Hierarchical Classification of Training Needs

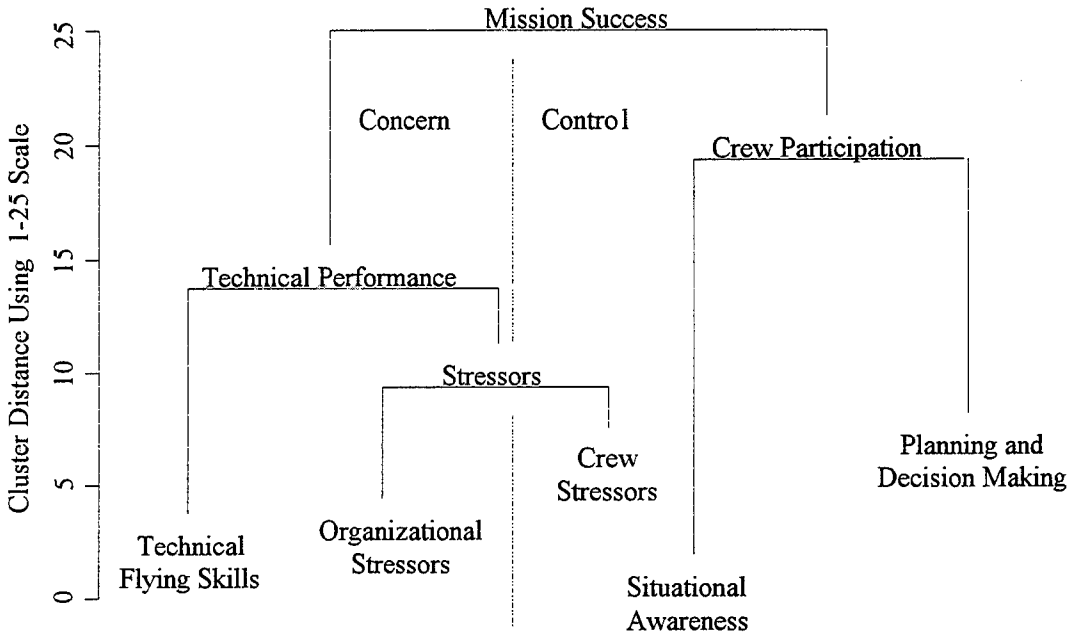


Figure 1 shows that mission success consists of two clusters, one that deals with Technical Performance, and one that deals with Crew Participation. Technical performance is further divided into issues related to technical flying skills as well as stressors that act to interfere with the performance of those skills. This interference consists of factors residing within the organization

as well as factors that reside within the crews. Crew participation is comprised of issues related to maintaining situational awareness as well as planning and decision making.

Discussion

The cluster analytic results of the training needs analysis support earlier findings that flight crew performance is dependent on three factors: (1) technical skills; (2) resource management skills; and (3) the organizational context in which flight crews operate. As shown in figure 1, technical performance and crew participation form two distinct classifications of training needs with organizational and crew contextual factors acting as stressors that interfere with the technical performance of flight inspection crews. Furthermore, the training needs that emerge from this classification may be further divided into two categories: (1) those factors over which flight crews have control; and (2) those factors that concern flight crews but whose control resides within the organiztion. This distinction is shown in figure 1 by the dashed line. Using the structural framework of figure 1, several training implications are especially worth noting.

First, crew members reported problems with the technical training they received. In particular, pilots complained that some of them were not getting enough flying time which made them feel not as technically proficient as they would have liked. In addition, pilots were not always checked out on equipment modifications prior to conducting a flight inspection mission. Since the single most important resource that crew members possess is the technical skills that they have acquired over time, technical training deficiencies such as these must first be addressed before crew resource management training can be expected to have a positive effect on crew performance.

Second, the results of the needs analysis suggested that crews would benefit by more active crew participation, particularly with regard to three areas: (1) pre mission briefings; (2) decisions about safety; and (3) maintaining aircraft situational awareness. The importance of a pre-mission briefing cannot be over emphasized. It is during the briefing that crews develop what Cannon-Bowers, Salas, & Converse (1993) and others have called a shared mental model of the mission. A shared mental model may be thought of as a common set of expectations of what will occur during the course of a mission. Included in this mental model are expectations concerning the time sequencing of mission events, the tasks to be performed, and how individual efforts will be coordinated. When a pre-mission briefing is lacking, crew members must rely on past experiences as a means to guide their performance. Because the specifics of the mission have not been communicated, crew members assume that everyone is operating with the same set of expectations. Unfortunately, it is usually under non-routine conditions that the fallacy of this assumption surfaces.

In addition to establishing a shared mental model of the mission, the pre-mission briefing is an excellent time to address crew stressors such as leadership, communications, and crew climate concerns. How a PIC conducts a pre-mission briefing sets the stage for the communication patterns that will emerge among crew members (Hackman, 1993). If the PIC provides a well organized briefing and solicits input for others, then he or she establishes an atmosphere of professional competency in which crew members feel free to voice their concerns. Furthermore,

to the extent that crews can resolve differences of opinions prior to flight, they are less likely to be distracted by those differences during the course of the mission. By involving all three crew members (PIC, SIC and ET) in decisions regarding flight safety, crews create a climate in which flight safety is a shared responsibility.

A third implication from the needs analysis concerns the effects that organizational stressors have on the technical performance of a flight inspection mission. Flight inspection crews are mission oriented Their job is to certify that a given facility's navigational aides are operating according to standard specifications. Due to a variety of reasons (such as a facility outage at O'Hare International, a high density traffic airport) there can be increasing pressure on flight crews to perform flight checks during marginal weather or during off peak traffic hours late at night. Job related stress can arise when flight crews perceive (correctly or incorrectly) that their management is more concerned about getting the job done then they are about flight crew safety.

Concerns about organizational stressors are valid and need to be addressed by the organization, however, caution is advised when addressing those issues during CRM awareness training. The inclusion of organizational factors is likely to shift the focus of CRM training from what Covey (1994) calls "areas of personal control" to "areas of personal concern." As shown in figure 1, contained in areas of personal concern are those issues that concern flight inspection crews, such as technical skills and organizational stressors, but whose control over instituting changes resides within the organization. In contrast, areas of personal control, such as crew stressors, situational awareness and planning and decision making, are more strongly associated with factors that crew members themselves have the power to change. Covey notes that there is a tendency for people to spend a considerable amount of time attempting to address areas of personal concern to the neglect of addressing areas of personal control. Because of this tendency, once organizational stressors are raised, CRM trainers may find it difficult to re-focus discussions on factors over which flight crews have personal control.

Finally, the results of the training needs analysis should be viewed from a broader perspective than just CRM awareness training. Crew Resource Management is more than a course, it is a philosophy that governs crew members', thoughts, feeling and behaviors during the course of a mission. Although basic CRM skills can be developed in a course, they are likely to fade over time unless awareness training is followed by: (1) annual CRM recurrency training; and (2) incorporating CRM principles throughout all levels of the organization (FAA, 1995). Whereas the former provides practice and feedback for CRM skill development, the latter provides the organizational reinforcement necessary to produce a lasting cultural change. Thus, for CRM training to be effective, an organization must be committed to a long term program of change. The issues raised in this report provide a starting point for beginning that process.

References

Cannon-Bowers, J., Salas, E., & Converse, S. (1993). Shared mental models in decision making. In N. Castellan (Ed.), Individual and group decision making. Hilsdale, NJ: Lawrence Erlbaum Associates, pp. 221-246.

Covey, S. (1989). Seven habits of highly effective people: Restoring the character ethic. New York: Simon & Schuster.

Federal Aviation Administration (1995). Crew resource management training (Advisory Circular 120-51B). Washington DC: author.

Hackman, J. (1993). Teams, leaders, and organizations: New directions for crew oriented flight training. In E. Wiener, B. Kanki, & R. Helmreich (Eds). Cockpit resource management. San Diego: Academic Press, pp. 47-70.

Hartel, C. & Hartel, G (1995). Controller resource management - What can we learn from aircrews (DOT/FAA/AM-95/21). Springfiled: VA: National Technical Information Service.

National Transportation Safety Board (1994). Aircraft accident report: Controlled flight into terrain, Federal Aviation Administration, Beech super king air 300/F, N82, Front Royal Virginia,October 26, 1993 (NTIS PB94-910405).

Towards a Unified Theory of Airmanship: A Model for Education

by
Tony Kern
United States Air Force Academy
Department of History
and
J.D. Garvin
United States Air Force Academy
Department of Behavioral Sciences and Leadership

Introduction

The military continues to experience tragic and embarrassing failures of airmanship, such as the Blackhawk shootdown and the B-52 crash at Fairchild AFB in 1994. These human failures damage military airpower credibility and warrants new investigation into air discipline and airmanship education. The purpose of this research was to define, conceptualize, and communicate a commonly-held structure and standards of good airmanship, with the goal to reduce or eliminate such tragic errors in the future. The researchers set out to accomplish three tasks, two of which are complete. First, the research sought to established a unifying definition (perhaps the first) of airmanship based on historical research into traits and characteristics of successful airmen and operations. Second, it expanded this definition to establish a conceptual taxonomy integrating its multiple constructs, creating -- The Airmanship Model. Finally, the research illustrates each airmanship construct and their overlapping integration through an educational media of case studies. When compiled, these case studies will provide a comprehensive, real world example of the previously ambiguous concept of airmanship. The Airmanship Model and case study collection will provide a structure upon which to hang a lifetime of learning, and subsequently develop the Tenets of Airmanship, or principles that will become a part of a more efficient, effective, and safe military flying culture of tomorrow.

Significance of the Study

Many approaches to airmanship education have been attempted. Most recently, a combination of human factors education and skill development combine the efforts of civilian Crew Resource Management (CRM) contractors and military flight trainers. This top down approach, mandated by AFI 3622-43 Cockpit/Crew Resource Management, has experienced many problems in implementation, including quality and flexibility in education and training, scheduling, and lack of an individualized approach in some of the larger CRM courses.

Additionally, recent high profile events highlight the continuing problems of poor airmanship. Consider the following incidents.

1. Two F-15 pilots, under the control of an American AWACS, misidentify, fire upon, and destroy two friendly helicopters resulting in an international incident. The wingman pilot lamented, "Human error did occur. . . It was a tragic and a fatal mistake which will never leave my thoughts, which will rob me of peace for time eternal. I can only pray the dead and the living find it in their hearts and their souls to forgive me." Further details are even more disturbing. Rules of engagement may not have been clearly understood, communicated, or followed.

2. A B-52 bomber crashes while executing prohibited maneuvers at a U.S. Air Force base. The investigation reveals that a rogue aviator had been allowed to consistently violate Federal Aviation and military regulations for at least three years. Even worse, this same aviator was the Chief of Standardization and Evaluation of all aircrew members in the wing. A minimum of five wing and operations group commanders had the opportunity to intervene during this time period.

3. Two A-10 pilots who were flying a close air support mission during DESERT STORM, misidentify British Warrior armored vehicles as an Iraqi armored column. They fire Maverick missiles into the allied vehicles, killing nine and wounding eleven British soldiers. A five-month British investigation into the incident attributed "no blame or responsibility to British forces." The British media splash the incident across tabloid headlines for months afterward (Powell, 1991).

4. An F-16, commanded by an experienced fighter pilot was on the first leg of a routine ferry flight for military sale flight to a foreign country. The fully functional aircraft never made it, as the pilot ran out of fuel and the aircraft crashed in route to a divert base (AETC, 1994).

5. An tower controller calls conflicting traffic "on short final" to an F-16 pilot conducting a simulated emergency approach well outside of prescribed operational guidelines. Although the pilot is unable to identify the traffic, he elects to continue the approach, resulting in a mid-air collision and the deaths of 24 Army personnel who are struck by the burning wreckage as they wait to board a C-141 for training (Cross, 1994).

These examples are just the surface symptoms of a very serious threat to military airpower. Shrinking military resources and heightened public awareness demand a more focused approach to the previously ambiguous concept of airmanship.

Methods

The researchers conducted an 18-month qualitative analysis of successful aviators of the past and present, seeking to gain insights on desirable traits of airmanship. One hundred fifty six aviators were identified for analysis, using a combination of qualitative analysis techniques including open, axial, and selective coding (Strauss, 1987). The researchers used a combination of primary and secondary research to construct their findings. In addition to the standard literature review on the traits of successful airmen, archival documents, personal papers, and interview transcripts were qualitatively coded and analyzed, data were then reconstructed, unveiling common themes of good airmanship.

Results

Defining Airmanship. When asked to define good airmanship, most aviators have difficulty. "I know it when I see it." is the response most often given (Kern, 1995). Words like judgment, discipline, and situational awareness, are often used, but few seem to be comfortable with an exact definition. The origins of this definition began as historical research from Operation DESERT SHIELD/STORM, a study that indicated tactical aircrew error had significant operational, as well as safety and training implications (Kern, 1994). During this study of error, the accidents and incidents begin to yield common themes of success. These themes took two forms: Common themes of what could have prevented the accident or incident, as well as successful behaviors leading to positive outcomes. After 18 months of investigation, our listing of desired characteristics evolved into a unifying, comprehensive definition of airmanship. Airmanship is the consistent use of good judgment and well developed skills to accomplish mission objectives. This consistency is built upon a cornerstone of uncompromising flight discipline, skills, and proficiency. A high state of situational awareness completes the airmanship picture and is obtained through knowledge of one's self, team, aircraft, environment, and the risk. Airmanship is seen as the measuring stick of the professional aviator. It is our professional obligation and intention to further define and develop behavioral characteristics that characterize ideal airmanship, as a model for self improvement for all. This can best be accomplished by meticulously defining and illustrating each construct of the new definition of airmanship or, in other words, developing accepted and measurable standards of airmanship.

Developing the Airmanship Model. Success leaves clues. Historically, successful aviators tend to possess certain common qualities and characteristics and a glimpse into the crystal ball of future technology or potential enemies suggests little change (Rippon & Mannel, 1918). The changes that have occurred overtime appear to be changes of degree only, and not fundamental shifts in the nature of what constitutes superior airmanship. This analysis revealed three fundamental principles of expert airmanship, regardless of the time frame analyzed; skill, proficiency, and the discipline to apply them in a safe and effective manner. Beyond these basic principles, five areas of expertise were identified as common among expert airmen. They are the knowledge of yourself, your aircraft, your risk, your team, and the environment -- both physical and regulatory. The model at Figure 1 illustrates the concepts uncovered by the research project.

The research suggests that an expert aviator combines various skills and knowledge into a comprehensive whole. A flyer with the right stuff is one who knows the capabilities and limitations of his aircraft, the environment, the risk, his teammates, and himself, and understands that all of these factors are dynamic -- requiring constant and calculated attention. An expert flyer builds upon a bedrock of flight discipline, skills, and proficiency. No single-focus flyer approaches excellence. A tactics expert who can't fly the aircraft effectively due to lack of proficiency, doesn't add much to the combat power of his country. Conversely, a golden hands pilot who doesn't understand the rules of engagement, or who misidentifies friendlies as foes, can do tremendous damage to his country's cause with a single error.

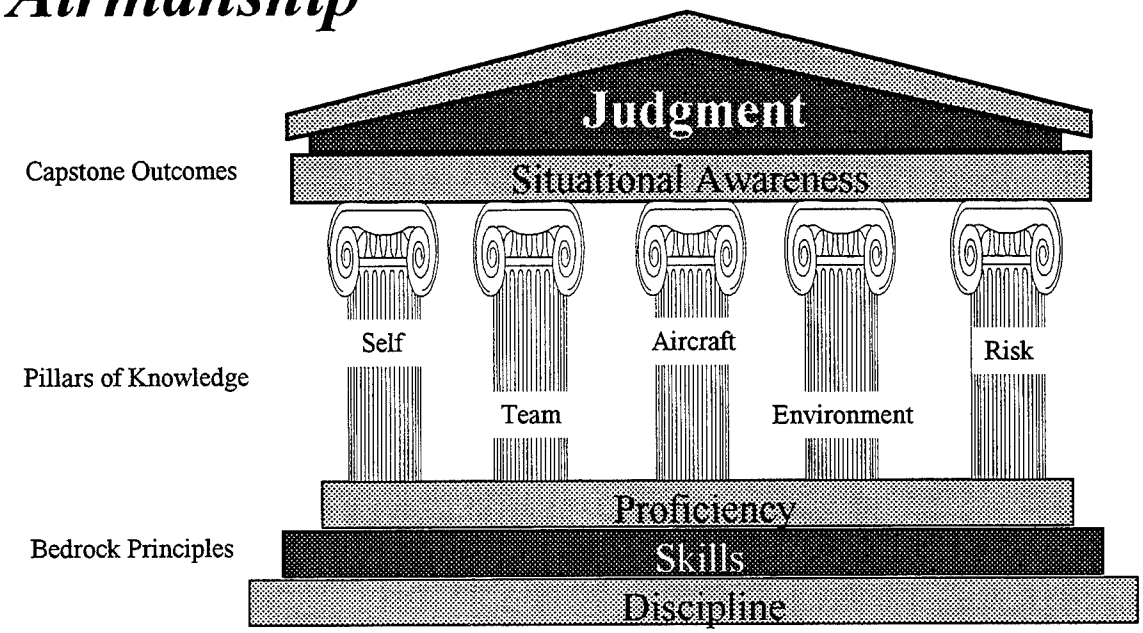


Figure 1. The Airmanship Model

The combination of tactical and technical expertise is still not enough. Even if the airman understands enemy systems and tactics, and can outfly everybody in the squadron, he is ineffective on the airland battlefield if he cannot integrate with his wingmen, crew, or the joint and combined team. This requires a special set of skills that have come to be known as human factors. The research indicates that total airmanship blends technical and tactical expertise, proficiency, and a variety of human factors to smoothly and effectively integrate the capabilities of the man and the machine into the joint/combined team. Total airmanship leads to improved situational awareness, fewer mistakes, increased operational effectiveness, improved training, and safer flying operations. By eliminating gaps in airmanship, a flyer is better able to handle the rapidly changing and dynamic environment of flight. But developing total airmanship is not a simple learning task. Reaching this level of expertise must start from within and begin with a motivation to improve--to develop an understanding of the skills and knowledge that will be required to carry the day--a basic understanding of what has come to be called airmanship.

Applications for Education. From a macroscopic perspective, the Airmanship Model may be useful in blending and phasing the overall aviation curriculum. Traditional models show the maturation of the aviator as a sequential, building-block process. The FAA model (Figure 2) builds towards a pinnacle of "Judgment" and is representative of many current training programs who stress layers of sequential training. This traditional "walk before you run" approach -- which

stresses learning the basics of flying before adding other types of training, has been used successfully in the military for decades. Nonetheless, failures of basic airmanship continue. A landmark study of incident and accident data by Foushee, reveals most problems of airmanship occur not because of a lack of proficiency or skill, but because of an inability to coordinate skills into effective course of action. Perhaps a multi-subject, integrated curriculum offers some potential for addressing the coordination and integration problems that continue to manifest themselves in poor airmanship in both the civilian and military sectors.

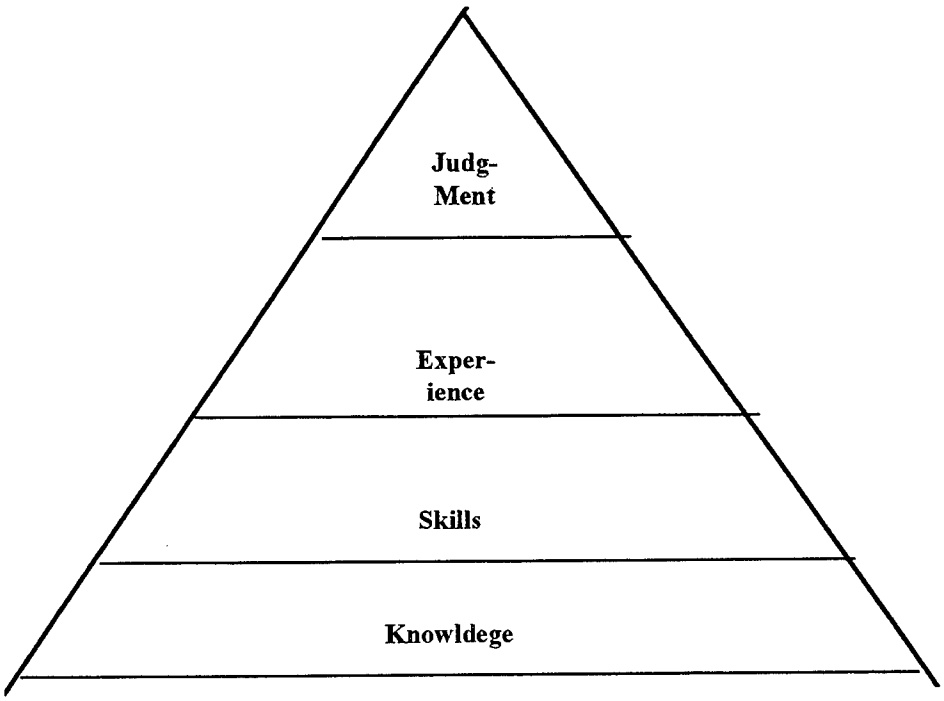


Figure 2. Traditional Layers of Aviation Training (FAA, 1987)

The airmanship model suggests an alternative approach, one that develops the aviator with parallel, rather than sequential design. Perhaps providing parallel instruction in knowledge of aircraft, environment, self, team, and risk -- would produce an aviator with a better sense for the integration of these various factors than the building block or layered approach currently used in many aviation training programs. A parallel approach could also avoid over-specialization or the development of "single-focus flyers" who excel in certain areas to the detriment or exclusion of others.

The airmanship model also presents new opportunities for utilizing case studies to show the integration of airmanship factors. The case study is best suited for teaching airmanship due to its real world application, its integration of all the curriculum components, and its active participation of the learner. The case study offers a real perspective of an airmanship scenario, complete with the ambiguities of an ill-defined problem, and the seamless integration of multiple components of the model. The student finds themselves in an active role in case study learning by having to define the problem or set of problems from the scenario, prioritizing decision making,

and implementing appropriate action. This is precisely how military pilots are currently trained for emergency procedures. Integrated airmanship education outside of this environment also promises positive results. The critical thinking skills and instinctive reaction should become common core to aviation paradigm. These complex behaviors are best developed through the real world illustrations that only case studies can provide.

Summary

Airmanship is clearly too important for relativistic interpretation. A common definition is the first step down the road to better understanding, educating, and implementing improved standards of airmanship. The potential benefits of an integrated systems approach include not only safety, but operational effectiveness and efficiency. The airmanship model provides a structure and relevance to operational and human factors education and training that is currently lacking. Today's aviators are victims of a disaggregated and fragmented approach to airmanship which separates and overspecializes to excess, leaving it up to the aviator to integrate various training and education, without ever being given a construct with which to do so. The airmanship model provides a means to build an integrated picture within the minds of many flyers -- the only place it really counts.

References

AETC (1994). Ignoring the Pinch. Torch, September 1994.

C. Cross (personal communication, September 8, 1994)

Kern, A. (1994). A Historical Analysis of Tactical Aircrew Error on Operations DESERT SHIELD/Storm. U.S. Army CGSC Monograph, 2 June 1994.

Kern, A. (1995). What is Airmanship? A survey of Military Aviators. Unpublished research in progress, United States Air Force Academy.

Powell, S. (1991). Friendly Fire. Air Force Magazine, December, 1991, p. 59.

Rippon, T.S. & Mannell, E.G. (1918). Report of the Essential Characteristics of Successful and Unsuccessful Aviators. The Lancet, 28 Sep 1918, pp. 411-418.

Strauss, A. L. (1987) Qualitative analysis for the social sciences. Cambridge: Cambridge Press

U.S. Air Force. (1994). Report on B-52 Mishap (AFR 110-14, 24 June 94).

U.S. Air Force. (1994). Report on Blackhawk Shootdown (AFR 110-14, Vol 12, p. 13).

An Evaluation of Full Flight Simulators and Flight Training Devices in Air Carrier Initial Flight Training Programs

John Wolf, Gerald Gibb, Steven Hampton, and John A. Wise

Embry-Riddle Aeronautical University, Daytona Beach, Fl

Abstract

The effectiveness of motion in flight simulators used to train and certify pilots is examined. Two groups of pilots were put through two similar training programs: one was a traditional program in which a full flight simulator (FFS) - including motion - was used for all training and certification, and the other program in which a FFS was used only at the final stage to measure pilot skill. This second group (the experimental group) received training in a simulator featuring all of the FFS features (including visual simulation) except motion. Training effectiveness was measured by using the simulator computer to determine the error in pilot control for six flight maneuvers. A flight instructor rating sheet was also filled out by an independent observer pilot.

Results show no significant differences between the training methods for four of the six maneuvers. In one maneuver, the angle of bank portion of the steep turn maneuver, the control group did perform significantly better than the experimental group. In the Visual approach maneuver, however, the experimental group performed better than the control group.

Introduction

Flight simulators have reached a high degree of realism in the presentation of visual, audio, tactile and motion cues. The most realistic simulators - full flight simulators (FFSs) - include out-the-window vision systems and motion platforms. Because of the proven training effectiveness of these simulators, aviation authorities have allowed the training and certification of pilots to be performed in these devices rather than the actual aircraft. The advantages to air carriers are lower cost when compared to aircraft training, and the ability to simulate maneuvers which, for safety reasons, would not be possible in the actual aircraft. The cost savings derive not only from the lower hourly cost of the simulator vs. the aircraft, but also from the increased training efficiency (it is possible to go directly to the desired location in space and practice the desired maneuver over and over without concern for the logistical constraints associated with real aircraft).

As the technical capability of simulators has increased over the years, the FAA has allowed and even encouraged their use in pilot training. In 1980, the FAA Advanced Simulator Plan (ASP) allowed the use of simulators for the final stages of training and checking. The ASP allowed the airline industry to further expand the use of simulators in training (Boothe, 1989). The ASP contained standards for three levels of simulators (Phase I through III), which when

added to the visual and non-visual simulator levels that previously existed, resulted in five levels of technical sophistication for flight simulators with Phase III being the highest, most realistic level. The intent of this structure was to allow the airlines to use the lower-level devices for lower-level training, preserving time in the more advanced devices for the most advanced training. From the regulatory point of view, the benefits of the program have been an elimination of training accidents and a much improved training environment. The benefit to the airlines has been the lower training cost of simulators when compared to the use of aircraft. However, the actual value of motion to training effectiveness has been the subject of many studies including Koonce (1974); see Waag (1981) for a review.

In 1993, Atlantis Aerospace Corporation performed a detailed analysis of the possible uses of various levels of simulators and trainers in a transition training program. As indicated in the Atlantis study, the objective of the demonstration was to determine whether pilot training costs could be reduced while maintaining the integrity and quality of the training program. Intrinsic to this purpose is the requirement not to compromise training effectiveness or certification standards and not adversely affect aviation safety. There can be more efficient flight simulator use by doing only necessary tasks in the simulator and doing other tasks in less costly devices. The key is to assign each task or event to the device which provides the necessary cues and environment for that task, but not to train in a more sophisticated device than necessary. It must be recognized that it is not sufficient for the pilot to merely accomplish the task. He or she must accomplish the task with the same control strategy and similar control inputs to those that would be used in the respective aircraft. The objective is for the flight training device (FTD) or simulator to provide the same pilot stimulus for the task that the aircraft would provide.

The need for simulator motion can be based on so called "disturbance" inputs which derive from unusual events or disturbances of the flight path as opposed to pilot-induced deviation from the flight path. Events which are known to be independent of motion stimulus can be trained, and for that matter checked, in an FTD. Training for some of the events may benefit from a visual system. In this demonstration, the visual system was used throughout the flight training portion of the program. There was no intent to identify which tasks would benefit by visual cues and which would not.

Current FTD and simulator task assignment is based primarily on realism. The issue is often how realistically does the device represent the total environment, not just how realistically it represents the given task or event. Realism is certainly an acceptable criterion for success, but it may lead to over-specification of the needed training medium. However, since there is no data base except experience that indicates what cues are required for given tasks, there is as yet, no other criteria. The demonstration described in this paper does not attempt to relate pilot response to cues, per se, but is shows that many tasks can be off loaded from the simulator to a less complex device. Hopefully, the results will stimulate further study into cue analysis.

Method

Purpose

The purpose of the research was to identify portions of pilot training that could be effectively conducted in a FTD with a visual system and which require the use of a FFS.

Subjects

A total of forty-eight pilots, twenty-four volunteer pilots from Embry-Riddle Aeronautical University and twenty-four volunteer furloughed pilots from Delta Air Lines, participated. In each case, half the pilots were in a control group and the other half in a test group. The crew training concept was used and all pilots in the captain position possessed, or were eligible for, an Airline Transport Pilot (ATP) certificate.

The mean flight experience of the Embry-Riddle pilots was 1300 hours with a range of 800 to 10,000 hours. Their mean age was 26 with a range of 22 to 43 years of age. Each pilot held at least a commercial certificate with a multi-engine and instrument rating. The Embry-Riddle pilots were paired so as to avoid having two low-experience pilots together as a crew, neither of which might possess the experience requirements for an air transport rating. After pairing, the subjects were assigned randomly to the control or test groups.

The Delta subjects were furloughed pilots of varying experience all of whom had previously served in a line capacity. They too were paired and then randomly assigned to the control or test groups.

All objective flight performance data were collected using the data collection capabilities of the simulator. All subjects completed the normal ten day MD-88 ground school which was an integral part of the initial training program. The ground training program was unaltered for the demonstration program and utilized the Level 6 FTD, but did not use the visual system.

The pilots in the control groups received flight training in accordance with Delta Air Lines standard all simulator (Level D) initial training program. The pilots in the test groups were trained in a program which used the visual FTD in lieu of the simulator for the first nine training and certification days in the program. Some tasks which require a simulator were learned and practiced in the FTD, but were then repeated in the simulator in the latter part of training program.

All subjects completed the check ride in the MD-88 flight simulator and were evaluated using standard performance criteria required by the FAA-approved Delta Airlines training program. The check rides were administered by an aircrew program designee (APD) who did not know whether the pilot was trained in the all simulator program or in the combined FTD and simulator program. Pass or fail was determined solely by the APD. Any pilot trainee needing more than the allotted time of the training program was given one additional day of training and a second check ride.

A second observer from the Embry-Riddle staff was present during the checking. His sole function was data collection. The observer, a senior check pilot, completed a detailed special performance evaluation form during each check ride. The analysis of the data from the special performance evaluation complemented objective data collected using the simulator computer system. The second observer also managed the collection of the objective data. This involved initializing the computer for data collection before each maneuver.

Results

The data from the Embry-Riddle independent observer consisted of rating sheets with simple dichotomous scores. Pilot performance was evaluated only as to whether or not a procedure, checklist item, or performance item was successfully completed within the parameters of the ATP practical test standards, which are identical to the performance required on a rating ride. Rating sheets were used to assess eight flight maneuvers: a) precision approaches, b) visual approaches, c) approach to stalls, d) non-precision approaches, e) normal takeoffs, f) rejected takeoffs, g) V_1 cuts and, h) steep turns.

The frequency of missed items was too small to analyze each of the eight maneuvers independently. Consequently, these data were combined across the maneuvers to develop composite ratings. Non-parametric tests were performed on these data between training conditions for the Embry-Riddle and Delta pilots separately, and combined as larger test and control groups. No differences between control and experimental training conditions were found for Embry-Riddle pilots (X^2=.14, ns), Delta pilots (X^2=0, ns), or the pilot groups combined (X^2=.08, ns).

Simulator-generated data was used in this study as a means to objectively assess and quantify performance while mitigating evaluator biases. The maneuvers and performance parameters were selected based on meeting three criteria: a) the maneuver was a required task in the checkride, b) performance could be assessed using captured relevant parameters, and c) a clear standard of target performance could be developed and used for comparison. All required tasks in the checkride could not be assessed since a clear reference point could not be obtained or because of difficulty in identifying the initialization or termination point. Therefore, only maneuvers and parameters that could be precisely standardized across all checkrides were used. No attempt was made to sample all checkride maneuvers or their components. The six maneuvers sampled and their associated performance parameters are described below.
The important performance criteria for this study, however, were the simulator captured data. Six sampled maneuvers were analyzed to determine if there were significant differences among pilots in critical flight performance measures. In each case, Group 1 represents the control group (full flight simulator throughout training) and Group 2 represents the experimental group (combined FTD and simulator). The reader is cautioned that complete performance data is not available in many instances as a result of simulator problems in capturing and transferring data. All analyses are conducted assuming unequal sample variances using the probabilities for two-tailed tests.

The data reported below have been organized by maneuver. In all cases where there are no significant differences between Embry-Riddle and Delta pilots within training conditions (i.e. control and experimental), the data have been collapsed to increase sample size. Results of these comparisons are not presented here. rot mean square (RMS) values were obtained by squaring each deviation value, summing the squares, dividing by the number of samples, and taking the square root of the result for each individual's performance. No distinction was made between first officers and captains.

Steep Turns:

The RMS of the deviations from 45 degrees angle of bank (AOB) from initialization (20 degree heading change from initial direction) to completion (within 20 degrees of final heading). Altitude and airspeed RMS deviations were acquired for the entire turn. Target airspeed and altitude is based on nominal target values at the initiation of the maneuver.

Rejected takeoff:

Root mean square of heading deviation from runway heading and the total distance to stop in feet. Data collection is initialized at loss of one engine (N2 reverses) and completion is at zero ground speed.

Engine failure at V1:

Root mean square of heading deviation from engine failure at V1 to restart.

ILS approach:

Root mean square of glide slope and localizer deviations in feet from five miles inbound to touchdown.

Approach to stall:

Mean number of feet of altitude lost between stall onset (yoke shaker flag) to recovery (increase in altitude after stall including any secondary stalls). Subsequent secondary stalls were treated as a continuation of the original stall.

Visual approach:

Distance from runway centerline at the point of touchdown.

	Group			Group					
	n	M	SD	n	M	SD	t	df	p
Steep Turns:									
Angle of Bank	20	1.86	0.68	25	3.28	2.91	2.35	27	0.02

(Table cont)									
Airspeed Dev	10	4.41	1.59	15	4.41	3.14	nil	22	ns
Altitude Dev	10	75.26	82.9	15	87.50	79.0	0.37	19	ns
Rejected Takeoff:									
Heading Dev	9	1.87	0.87	13	1.76	1.19	0.25		ns
Distance to stop	8	1113.9	364.2	7	1320.5	559.8	0.83		ns
		Group			Group				
	n	M	SD	n	M	SD	t	df	p
Engine Failure:									
Heading Dev	15	4.26	1.60	16	3.67	1.79	0.97	28	ns
ILS Approach:									
Horizontal Dev	21	0.53	1.94	24	0.11	0.05	1.00	20	ns
Vertical Dev	21	1.39	0.89	24	0.97	0.81	1.65	41	ns
Approach to Stall:									
Altitude lost	19	80.7	73.1	16	52.8	52.2	1.33	32	ns
Visual Approach:									
Embry-Riddle	9	15.19	14.70	18	23.76	28.70	1.03	25	ns
Delta	20	14.05	12.15	17	6.85	5.45	2.38	27	0.024

Conclusions

Only six maneuvers were evaluated in this study. However, maneuvers were selected which could be objectively measured and evaluated. Other maneuvers and training tasks were not evaluated.

In four of the six maneuvers evaluated no significant differences were found between the control and experimental groups. In one of the maneuvers, Steep Turns, the control group outperformed the experimental group. In the remaining maneuver, Visual Approach, the experimental group was slightly better than the control group. Data gathered by the second flight instructor/observer also showed no significant difference between the groups.

The results of this study lend support to the concept of transferring some of the flight training tasks to devices which are less complex and less costly than full flight simulators. Other tasks, for example steep turns, may benefit from the added realism of motion in simulation.

References

Boothe, E.M., Cook, E.D. (1989) FAA perspective in increasing benefits of flight simulation.

Proceedings 1989 spring convention - flight simulation: Assessing the benefits and economics. London: The Royal Aeronautical Society.

Federal Aviation Administration (June, 1980) Federal Aviation Regulation Part 121, Appendix H Advance Simulation Plan.

Koonce, J.M. (1974) Effects of ground-based aircraft simulator motion conditions upon prediction of pilot proficiency. Savoy, Ill.: University of Illinois, Aviation Research Laboratory, TR ARL-74-5/AFOSR-74-3 (Ph.D. dissertation, University of Illinois at Urbana-Champaign).

Waag, W.L. (1981) Training effectiveness of visual and motion simulation. Williams AFB, AZ: AFHRL-TR-79-72.

Panel

Shaping Tomorrow's Military: The National Agenda and Youth Attitudes

Abstract

Here are several issues shaping tomorrow's military both on Capitol Hill and in the minds of the nation's youth. Pertaining to recruitment of youth to join the officer ranks, this panel will address the changing Congressional agenda, recent Congressional actions, the 104th Congress, and potential legislative programs and Defense initiatives that may emerge in 1997. The panel will also present a longitudinal look at youth attitudes toward the military, with a particular focus on their interest in officer training programs. We hope to show the impact and truth in the statement: "What happens in Washington affects the Nation and what happens in the Nation impacts upon legislative programs and policies in Washington."

Panel Members

Dr. W. S. Sellman
(Chair and Discussant)
Director for Accession Policy
Office of the Assistant Secretary of Defense
(Force Management Policy)

Mr. William J. Carr
(The National Agenda)
Assistant Director, Officer Accession Programs
Office of the Assistant Secretary of Defense
(Force Management Policy)

Major Dana H. Lindsley
(Youth Interest in College-Level Officer Training Programs)
Assistant Director, Recruiting Research and Analysis
Office of the Assistant Secretary of Defense
(Force Management Policy)

The National Agenda

A Changing Congressional Agenda...

This segment would review how an apparently changing ideology in Congress directly affects officer accession programs, as well as the future shape of the Department's legislative initiatives and policies in shaping its accession programs.

Recent Congressional Activities...

- Longer service obligations for academies
- Requirement for greater regulation
- Initial appointments must be in the Reserves
- Greater civilianization of academy faculties
- GAO involvement (11 recent academy reviews)
- Privatization study - academy prep schools
- Scrutiny of academy athletic programs

104th Congress...

- Shorter service obligations for academies
- Requirement for less regulation
- No new GAO reviews
- Repeal of academy prep school privatization study
- Repeal of laws governing academy athletic programs
- Financial penalty for schools hostile to ROTC

Potential Legislative Initiatives (FY 1997)...

- Slightly relaxed age standards -- academies and ROTC
- Montgomery GI Bill eligibility for some ROTC scholarship recipients
- ROTC scholarships for selected graduate school students

Potential Policy Reforms...

- Viability standards established for ROTC units (minimum 15 annual graduates)
- But political counterpressures limit closure of ROTC units and headquarters, leading to...
- Potentially too-large infrastructure (units no longer viable), taxing resources, leading to...
- Sharp reductions in the value of certain scholarships, leading to...
- Different attributes for scholarship recipients (e.g., ACT/SAT scores), perhaps leading to...
- Changes in officer performance and retention.

A Re-focused DoD Agenda...

The chain of events is leading not only to DoD proposed changes in its legislative program, but also to increased efforts to more-systematically capture data -- centralize storage of existing data elements, to help evaluate the impact of laws and policies on officer performance and retention. This, in turn, can help to improve legislative or policy actions, and resource allocations.

Interest in College-Level Officer Training Programs
Youth Attitude Tracking Study (YATS
1992-1995

Youth Attitude Tracking Study...

Since 1975, the Department of Defense annually has conducted the Youth Attitude Tracking Study (YATS), a computer-assisted telephone interview of a nationally representative sample of 10,000 young men and women. This survey provides information on the propensity, attitudes, and motivations of young people toward military service. Enlistment propensity is the percentage of youth who state they plan to "definitely" or "probably" enlist in the next few years. Research has shown that the expressed intentions of young men and women are strong predictors of enlistment behavior.

Trends...

Results from the 1995 YATS show propensity was slightly higher than in 1994. In 1995, 28 percent of 16-21 year-old men expressed positive propensity for at least one active-duty Service, up from 26 percent in 1994. Propensity for the Army and Navy also increased while propensity for the Marine Corps and the Air Force did not change. Propensity of 16-21 year-old women in 1995 was generally unchanged from 1994. However, 7 percent of 16-21 year-old women expressed propensity for the Air Force, a statistically significant increase from 5 percent in 1994, but the same level observed in 1992-93. Propensity among 22-24 year-old men and women was unchanged from 1994.

Summary...

Over the past several years, enlistment propensity has declined as the Services experienced serious cuts in recruiting resources. In 1994-95, recruitment advertising was increased, and the 1995 YATS results indicate that the decline in propensity may have abated. Continued investment in recruiting and advertising resources is required, however, to assure that the pool of young men and women interested in the military will be available to meet Service personnel requirements in the future.

Youth Interest in Officer Training Programs...

In the 1992, 1993, 1994, and 1995 YATS, a representative sample of American youth were asked about their attitudes toward becoming an officer in the military. Youth who responded to the YATS indicating they would like to complete at least four years of college (a Bachelor's degree) were asked about their interest in college officer-training programs. This discussion will address respondents interest in participating in a college program that would prepare them to become military officers, which type of program would be preferred (ROTC, OTS, Service Academy), which Service's program they would prefer (Army, Navy, Marine Corps, Air Force, Coast Guard), what information shaped their attitudes (family, mail, TV, a friend, etc.), reasons they would want to become an officer, as well as several other factors.

Determinants of Military Allied Health Care Students' Success: A Multifactorial Analysis

Captain Russell D. Porter, Ph.D.
Captain Jimmy L. Sterling, Ph.D. Candidate
Captain Joy P. Vroonland, Ph.D.
GS-14 Squy G. Wallace, Ph.D.

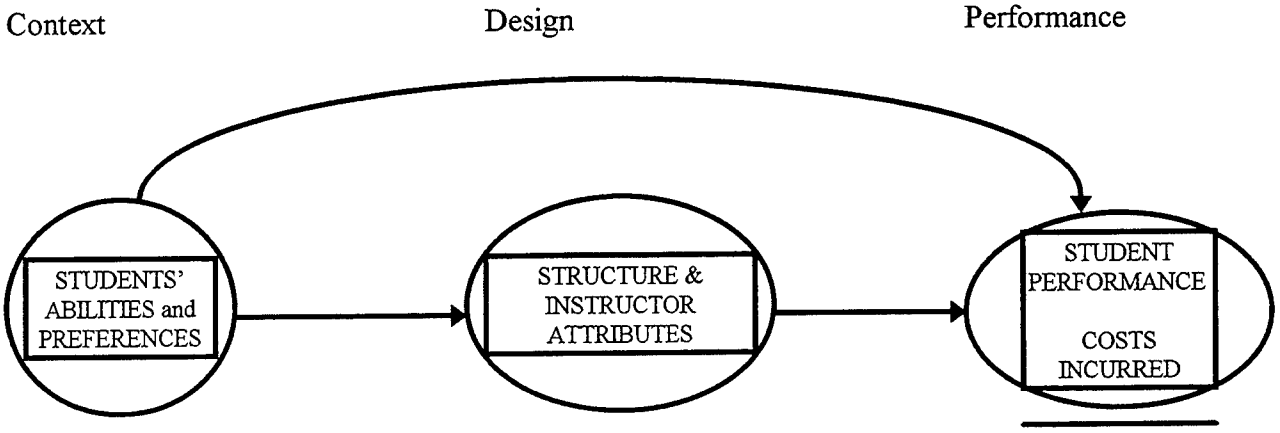
Abstract

Determinants of students' success in military and civilian allied health care training has traditionally focused on students' attributes. However, using the rational contingency theory as a framework, organizational and instructor attributes, as well as costs incurred, may significantly effect students' success as much or more than students' attributes.

This study will attempt to determine the degree to which student indicators, organization indicators, instructor indicators, and costs incurred, effect students' success. Specifically, indicators assessed will be: (1) students' standardized test results established prior to instruction (students' abilities and preferences), (2) structural indicators such as use of Learning Center interventions (organizational attributes), (3) instructors' education, experience and professional military education (instructors' attributes), and (4) manning (i.e., employee), supply, and overhead expenditures (costs incurred). Students' success is indicated by attrition rates during initial instruction, as well as ability to perform tasks during internships (i.e., phase II training).

The multifactorial technique will include traditional regression analysis and analysis of variance, along with confirmatory and structural procedures. Using student aggregate level results, recommendations will focus on improving recruitment requirements, organizational structure, instruction in the classroom, decreasing costs, and ultimately improving students' success.

Rational Contingency Framework



(Source: Kaluzny & Veney, 1980)

Panel Session:
Life Aboard a U.S. Aircraft Carrier: Examinations of Biomedical and Safety Issues.

Panel Chair:
Robert Stanny, Ph.D.
Naval Aerospace Medical Research Laboratory

Panel Presentations:

The Stress and Strain Associated with Deployment Aboard a U.S. Aircraft Carrier
Doug Wiegman, Ph.D., University of North Florida

The Stress, Strain, and Work/Rest Cycles of U.S. Navy Aircraft Carrier Flight Deck Personnel aboard a U.S. Aircraft Carrier
David McKay, Ph.D., Circadian Technologies &
LT Dylan Schmorrow, Ph.D., Naval Air Warfare Center - Aircraft Division

The Naval Flight Deck: An Unforgiving Environment for the Untrained or Complacent
LCDR Scott Shappell, Ph.D., COMNAVAIRLANT

Perceptions of Stress and Strain: An Examination of Flight Deck Crew Interviews
LT Dylan Schmorrow, Ph.D., Naval Air Warfare Center - Aircraft Division,
Claire Portman, Naval Aerospace Medical Research Laboratory, &
David McKay, Ph.D., Circadian Technologies

Naval aviation is an inherently dangerous and unforgiving environment. Research efforts to date have generally focused on the operator (aircrew) to minimize risks associated with naval aviation. However, as important as the aircrew are, a similar amount of research and information is sorely lacking regarding those who make it possible to fly the aircraft, the flight deck personnel. Operating in an equally unforgiving environment, these individuals are often asked to work extended hours, with variable opportunities to sleep. It is well known among officers in charge of flight decks and air operations that flight-deck personnel are among the most over-worked individuals aboard ship. There is no question that difficult work/rest schedules and an unforgiving naval flight deck combine to create a potentially hazardous environment. The purpose of this panel is to examining current research in this area and document these issues. This work represents the combined efforts of the Naval Aerospace Medical Research Laboratory, the Naval Air Warfare Center - Aircraft Division, and the Commander Naval Air Force Atlantic Fleet. Methods of analysis included: (1) an initial stress questionnaire; (2) daily activity logs; (3) flight schedules; (4) informal observations from daily interactions with ship's company; (5) structured interview data examining stress/strain and quality of life issues; and (4) accident data. Major stresses on flight deck and stresses outside of work are examined. Recommendations shall be discussed.

The Stress and Strain Associated with Deployment Aboard a U.S. Aircraft Carrier

Douglas A. Wiegmann, Ph.D.
University of North Florida

This portion of the panel session will focus on a recent field study designed to (a) identify the relationship between work stress and strain experienced by flight-deck personnel during deployment aboard a United States naval aircraft carrier, and (b) examine the potential role that social variables and diurnal type (i.e., morningness/eveningness) play in buffering the effects of stress on strain. Data were collected using a questionnaire that was completed by flight-deck personnel during the fourth month of a six-month deployment. Results of the study indicated that stress due to the working and living conditions aboard the carrier was significantly related to psychological and physiological strain and to the frequency of reported illnesses and accidents. Stressors tended to have an additive effect on strain, such that strain increased as the number of stressful events increased. Buffer variables were related negatively to strain; strain decreased as social support and "morningness" characteristics increased. Morningness had compensatory effects, reducing the psychological strain produced by stress sources. These findings suggest that improvements in the occupational health and well-being of flight-deck personnel could be accomplished by improving the living and working conditions aboard aircraft carriers. Some suggestions for improving these conditions will be discussed.

Work/Rest Cycles and Strain Among Flight Deck Personnel aboard a U.S. Aircraft Carrier.

David McKay, Ph.D.
Circadian Technologies
&
LT Dylan Schmorrow, Ph.D.
Naval Air Warfare Center - Aircraft Division

This portion of the panel session will focus on the work/rest cycles of U.S. Naval flight deck personnel aboard a deployed aircraft carrier. The purpose of this study was to examine the work/rest cycles and to assess the relationship between work/rest cycles and strain experienced by flightdeck personnel during a 6 month deployment in the Adriatic to support the United Nations in a peacekeeping mission, Operation Deny Flight. While aboard the U.S.S. Dwight D. Eisenhower (CVN-69), 146 flight deck personnel completed daily activity cards upon which they recorded when the worked, rested, ate, and exercised in 1/2 hour increments. Activity cards were collected for 72 days during the latter half of the mission. Results indicated that work/rest cycles varied daily depending on the type of mission being flown and the individuals particular work group. Catapult and arresting gear operators (CAT/AG) was the group that experienced the greatest variability in work/rest cycles. The relationship between cycle variability and strain experienced by flight deck personnel will be discussed.

The Naval Flight Deck: An Unforgiving Environment for the Untrained or Complacent

LCDR Scott Shappell, Ph.D.
COMNAVAIRLANT

Before addressing how to minimize hazards associated with naval flight decks, the hazards themselves must first be documented. This portion of the panel session will focus upon injuries sustained by personnel working on naval flight decks between January 1977 and December 1991. Data included all fatalities, permanent total disabilities, permanent partial disabilities, and major injuries resulting in five or more lost work days. A total of 918 flight deck personnel were reported injured during this 15 year period, including 43 fatalities, a plethora of fractures, traumatic amputations, major lacerations, dislocations, contusions, concussions, burns, crushing injuries, sprains, and strains. The most common, and arguably most complex, naval flight decks are located on the twelve active U.S. Navy aircraft carriers. As such, the vast majority of injuries occurred on these platforms. However, flight decks can be found on a variety of other platforms including amphibious, escort, and auxiliary ships. In fact, nearly all naval platforms with a flight deck reported a serious injury. An examination of the current injury rate aboard these naval platforms revealed an average of 51 serious injuries per 100,000 aircraft recoveries between 1977-1986 followed by a marked reduction to an annual rate of roughly 30 injuries per 100,000 aircraft recoveries between 1987-1990. What makes injuries sustained on the flight deck particularly disconcerting is that over 90 percent can be attributed to human causal factors.

Quality of Life and Perceptions of Stress and Strain: An Examination of Flight deck Crew Interviews

LT Dylan Schmorrow, Ph.D.
Naval Air Warfare Center - Aircraft Division

Claire Portman
Naval Aerospace Medical Research Laboratory
&
David McKay, Ph.D.
Circadian Technologies

This portion of the panel session will focus on data obtained through structured interviews with flight deck personnel aboard a deployed aircraft carrier. The purpose of this study was to gain insight into flight deck crew perceptions of their working and living environments. These interviews were conducted on board the U.S.S. Dwight D. Eisenhower (CVN-69) during a deployment in the Adriatic Sea in support of a United Nations peacekeeping mission. Interviews included enlisted flight deck workers from V1, V2, and V4 and officers in charge of flight deck operations. An overview of these interviews shall be presented and apparent trends will be identified. Major stresses identified included stress originating from supervisors, concerns for safety, and problems with coworkers. Major stresses outside of work included long lines, lack of privacy, and showering/berthing conditions. Discussion points to be addressed include management and supervisory education for dealing with maximizing personnel performance, training line supervisors to detect early signs of fatigue and with training flight deck personnel in sleeping habits, eating habits and stress reducing activities.

Team Effectiveness in the Space Launch Environment: Theory to Application

Jeffrey S. Austin, Ph.D
United States Air Force Academy
Robert C. Ginnett, Ph.D.
Center For Creative Leadership
Barbara G. Kanki, Ph.D.
Cheryl M. Irwin
NASA-Ames Research Center
Earl R. Nason, Ph.D.
United States Air Force Academy
Timothy S. Barth
Patrick S. Simpkins
NASA - Kennedy Space Center
Donna M. Blankmann-Alexander
Mark J. Nappi
Lockheed-Martin Space Operations

One of four major industrial engineering functions in support of Kennedy Space Center (KSC) Shuttle processing is methods engineering in which tasks are designed to minimize cost and worker effort while maximizing safety and quality. Methods engineering within the KSC "factories" includes consideration of several unique human factors issues. At KSC nearly all processing tasks are performed by teams rather than individual workers. A multi-organizational, multi-disciplinary team has been examining team effectiveness within the KSC complex for three years in the validation and use of Hackman's (1990) Team Effectiveness Model, later revised by Ginnett (1993). The purpose of this panel is to discuss some of the methods, results and applications of this major research effort. Ginnett will provide an overview of the model and how it has modified as a result of the research efforts. Austin will discuss the methodology and the creation of research teams capable of observational data collection in secure environments. One of the key elements of the model is group process. Kanki and Irwin will discuss an approach for the collection and analysis of team process data in ground maintenance operations including an example from the KSC environment. Nason will present results of a content analysis of the observational data. Three of the themes are discussed in terms of diversity in work teams. Finally, Barth, Simpkins, Blankmann-Alexander and Nappi will demonstrate one application of the model to help senior leaders track and understand the human factors implication of accidents and mishaps. This tool enables the assessment to focus on the human factor issues relevant to the work system, the physical system and the social system. The model allows a more in-depth analysis of causal factors. For example, the team will show how the model led to different conclusions about causes of previous mishaps and incidents. The tendency had been to look at technical fixes. The model has led to raising the awareness level of the impact of teams and team leadership on effectiveness.

Recent Developments in Methods for Formative Evaluation in Military Education

Winston R. Bennett, Jr., Ph.D.
USAF Armstrong Laboratory/HRT
Kent L. Gustafson, Ph.D.
William Wheeler
University of Georgia

Abstract

This part of the panel presents recent activities related to the development and application of methods for formative evaluation of a Civil Engineering course conducted by the US Air Force Academy. Formative evaluation involves using evaluation information to make changes in the structure and content of an education or training course or program. The development and application of several innovative methods for gathering formative evaluation information are discussed. These methods include: structured student diaries; field observations of the instructional and learning environment and process; and instructor questionnaires. Further, the integration of the information obtained from these methods as part of a comprehensive program evaluation is presented using two recent case studies related to the Civil Engineering course. In each case, the usefulness of the information is examined in terms of its role in formative evaluations of the course in terms of providing prescriptions for course revision, identifying exemplars of context-based and engaging situations, and obtaining periodic attitude and learning assessments of the students. The integrated approach described in this paper provided valuable information for formative evaluation and for substantive changes in the process and content of the course. Recommendations for the development and application of these methods for program evaluation are presented.

Predicting Crew Resource Management (CRM) Aspects of Aircraft Commander Performance Using a Situational Judgment Test

Kenneth T. Bruskiewicz
Jerry W. Hedge
Mary Ann Hanson
Kristi K. Logan
Walter C. Borman

Personnel Decisions Research Institutes, Inc.

Frederick M. Siem

Armstrong Laboratory Human Resources Directorate

Abstract

For decades, pilot selection in both the military and commercial sectors has focused primarily on the identification of individuals with superior flying skills and abilities. More recently, the aviation community has become increasingly aware that successful completion of a flight or mission requires not only flying skills, but also the ability to work well in a crew situation. In the current research, a CRM situational judgment test for Air force transport pilots was developed and validated. Situational judgment tests require respondents to read a series of job-relevant situations and then indicate which of several alternative actions would be most effective and which would be least effective in each situation. The current test was developed with the assistance of experienced aircrew personnel who generated brief descriptions of challenging and realistic interpersonal situations that aircraft commanders might face on the job, as well as a variety of viable response options for each situation. This effort resulted in a set of 60 difficult situations, and a representative sampling of the types of responses pilots might make in these various situations. This test, the Situational Test of Aircrew Response Styles (STARS), can be used to identify individuals likely to perform well in a crew situation. The test was validated using behavior-based performance rating scales targeting seven different aspects of CRM aircraft commander performance. Using these scales, aircraft commander performance ratings were collected from individuals in all crew positions at the same time that the test was administered. Validity coefficients are compared and contrasted across each crew position, and with self-ratings of aircraft commander performance.

Behavioral Sciences Career Field Review

William H. Cummings, III, Lt Col, USAF
HQ USAF/DPXET
Washington, D.C.

Abstract

HQ USAF/DPXET undertook an informal review of the 61SXB career field, in response to several problems: Low promotion rates for Behavioral Scientists, lack of clear career paths for junior Behavioral Scientists, utilization problems identified by organizations, and shrinking numbers of senior positions, due to downsizing and civilianization. The results of a formal HQUSAF/DPX Career Field review indicates that 61SXB is a viable career field, for now. Some of the findings include: Increased burden on senior Behavioral Scientists to solve problems from within, and Behavioral Scientists need to optimize their own career management. This panel/workshop will focus on findings from DPEXT's four lines of inquiry: List of behavioral Sciences positions and locations, highlighting "major users", career advice from senior Behavioral Scientists, commanders, and other experts, promotable/nonpromotable career paths, and preliminary results from 61SXB Occupational survey. Workshop will include information on key contacts and a question and answer session.

Educational Innovations: Advanced Technology Assessment

Todd A. Fore, M.A.
82d Training Group
Air Education and Training Command

Abstract

Advanced technology assessment is necessary to determine if specific training technologies are having a positive impact on improving trainees' knowledge or skill performance. This research: (1) reviews the many innovative classroom technologies that exist in resident and non-resident training within today's Air Education and Training Command; (2) presents the initial interactive customer-based (student) assessment of exposure to distance learning and computer based instruction; and (3) establishes the increasing need to conduct empirical research on the use of technologies in training. The paramount axis for the application of educational technology seems to be the conjunction of media and instructional methods which results in a greater amount of knowledge or skill performance.

A Leadership And Communication Skills Development Training Program for Airport Checkpoint Security Supervisors: Development and Evaluation

Gerald D. Gibb, Ph.D.
Sam C. Kelly, M.A.S.
James S. Baker, M.A.S.
Daniel Sola, M.A.S.
Colleen Wabiszewski
Xavier Simón
Embry-Riddle Aeronautical University

Abstract

The foundation for this research program is from a two-year study that examined the job satisfiers and dissatisfiers of airport security personnel nationwide. The results indicated a critical need for supervisor leadership and communication skills training. Further supporting evidence was obtained from an analysis of an organizational climate survey conducted within the industry. The impetus for this effort was dictated by a requirement to mitigate the severe personnel turnover rates in this environment.

In response to a need to establish standardized, portable training that will improve job performance and reduce personnel turnover at the nation's airport security checkpoints, a twelve hour curriculum was developed. The content areas for the training program were developed in light of previous findings and included training in leadership development, basic supervisory skills, communication techniques, and conflict resolution skills. The training program was implemented at major airports within the U.S.

Concurrent with the development of the training program, several operational performance criteria and parameters were defined that were related to the training content areas. Consequently efforts are underway toward the assessment of operational performance criteria to measure the impact of the training. The emphasis of these efforts is to quantify those performance elements that have operational significance. Results from the initial two-year study, the training program curriculum, and available evaluation data are presented.

Training Ship Handling Skills in a Virtual Environment: A Comprehensive Requirements Determination Approach

Robert T. Hays, Ph.D.
Rosemary Garris-Reif, M.S.
Naval Air Warfare Center Training Systems Division

Abstract

This poster session will describe the status of an ongoing Advanced Development (6.3) program called "Virtual Environment for Submarine Ship Handling and Piloting Training (VESUB)." The VESUB project is developing a technology demonstration that incorporates state-of-the-art virtual environment, head-mounted display, and instructional technologies to train submarine Officers of the Deck (OOD) to safely maneuver their ship into and out of port. The project will result in the first Navy training application of virtual environment technology. The VESUB effort will proceed through five major stages: 1) requirements determination; 2) formative system development; 3) system enhancements; 4) training effectiveness evaluations; and 5) development of system procurement specifications and recommendations for insertion into Navy training programs. Currently, VESUB is in the second stage of development. This poster describes the approach developed to ensure that the requirements for training the submarine ship handling task are effectively and efficiently articulated. This approach requires the interaction of five teams: 1) government researchers; 2) an implementation planning group of fleet subject matter experts; 3) the technology demonstration system development contractor; 4) a contractor who will help determine instructor/operator station, performance measurement, and system interface requirements; and 5) a submarine subject matter expert contractor. The results of the requirements determination phase and plans for follow-on development will be presented.

Effects of Aerobic Fitness and Individual Characteristics on Cardiovascular Reactivity and CHD Potential

William H. Hendrix
Clemson University
Richard L. Hughes
Center for Creative Leadership

Abstract

The purpose of this research was to provide a partial test of a cardiovascular reactivity (CVR) model which incorporated antecedents of CVR (aerobic fitness, Type A behavior, and Trait variables) and the resulting effects of these antecedent factors, on CVR and the risk of developing coronary heart disease (CHD). Recent research has suggested that "hot reactors" who display increased CVR under stress are more likely to develop coronary heart disease. Variables under investigation included aerobic fitness, trait anger, trait anxiety, two measures of Type A behavior, four blood pressure measures under stress, and the cholesterol ratio as a measure of CHD potential. Subjects consisted of 134 Department of Defense senior male military and civilian employees assigned to a senior service school. The data were collected as a part of a health promotion program where subjects were subjected to stress using a video game and blood pressure measures taken to assess the extent of CVR. Results indicated that the major contributors to CVR were Hard Driving Type A behavior which increased CVR and CHD potential. Aerobic fitness had a direct effect on CVR where higher levels of fitness where associated with lower CVR. CVR in the form of mean diastolic blood pressure under stress was directly related to increased CHD risk. A revised model of CVR was developed which more adequately depicts the CVR relationships than did the hypothesized model.

Content Analysis of Employees' Reports of Their Feedback Seeking Behavior

Ann M. Herd, Ph.D.
Capt Heather Pringle, M.S.
United States Air Force Academy

Recent empirical and conceptual research studies of employees' feedback seeking behavior suggest that employees may seek performance feedback using various strategies from a variety of sources and for a variety of reasons (Levy et al., 1995). For example, employees may seek feedback from their supervisor, peers, subordinates, and clients. They may directly inquire, or ask, the source for feedback, or they may monitor the source's reactions to their performance. Employees may seek feedback to help them improve their performance to achieve important organizational and individual goals, or they may seek feedback for "impression management" purposes, to present themselves to the feedback source in a positive light.

Studies of employees' feedback seeking behavior have largely used Likert-scaled, questionnaire measures to assess reported feedback seeking behavior as well as factors hypothesized to be associated with this behavior. In addition, no studies have directly assessed impression management as a reason for seeking feedback. The present study uses a content analysis procedure to investigate employees' reports of their own feedback seeking behavior under conditions of successful and less than successful performance. Subjects in the study were 107 graduate students in a master's-level Organizational Behavior course at a small northeastern college. Students in the course were given traditional questionnaire measures of their feedback seeking behavior, which they were asked to complete after recalling a specific example in the past of successful and unsuccessful performance. They were then asked to use the diagnostic approach (Gordon, 1994) to describe, diagnose, and prescribe alternatives for their behavior in these feedback seeking situations. These reported incidents will be analyzed with regard to the following factors: 1) the reason for seeking feedback; 2) source and strategy of the feedback seeking attempts; and 3) environmental and personal factors relating to the reason, strategy, and source of feedback seeking. The frequency of strategies, sources, and reasons for seeking feedback will be reported. In addition, impression management as a reason for seeking feedback, and the factors relating to impression management feedback seeking efforts, will be specifically reported. Implications of the findings will be discussed in light of current research on feedback seeking behavior as well as avenues for future study.

References

Gordon, J.R. (1993). A diagnostic approach to organizational behavior. Boston, MA: Allyn & Bacon.

Levy, P.E., Albright, M.C., Cawley, B.D., & Williams, J.R. (1995). Situational and individual determinants of feedback seeking: A closer look at the process. Organizational Behavior and Human Decision Processes, 62(1), 23-37.

Low-Visibility Surface Operations: Crew Navigation Strategies and Use of Taxi Maps

Cheryl M. Irwin, M.A.
Kim E. Walter, B.A.
San Jose State University Foundation
NASA Ames Research Center

Abstract

Adverse weather conditions can put considerable strain on the National Airspace System. As visibility approaches landing minima at airports, operations grind to a halt. Even small decreases in visibility on the airport surface can create delays, hinder safe movement and lead to errors. This study analyzed data from a 747 simulation study evaluating the use of a moving map display for ground taxi in low visibility. Twelve two-person crews each conducted ground taxi trails in VFR, 600 ft. or 300 ft. visibility conditions. A crew was assigned to one of three map conditions. The first condition used the traditional paper map. The second used the basic moving map, which provided taxi way information and aircraft position. The final map condition, the advanced moving map, provided other traffic and a highlighted route to the gate in addition to the basic map information. Low-visibility disorientation errors were identified and coded for contributing factors, communications preceding the error and consequences of the error such as delays. We examined the effectiveness of crew strategies for detecting and recovering from errors, taking into account map usage and coordination of navigation and orientation information. Crews using the moving map had fewer errors overall and were able to detect errors and recover from them more quickly than crews using the traditional map. We compared crew communication strategies for dealing with complex taxi way configurations to determine which information is pertinent and what level of detail is appropriate to effectively taxi in reduced visibility. Recommendations for coordination of information and crew strategies for taxiing in low visibility are discussed.

Using Gender Role Conflict Scores to Predict Requests for Different Types of Counseling Services

R. Jeffrey Jackson
Christopher R. Kieling
Jonathon R. Eckerman

USAF Academy

Abstract

Gender has an organizing effect on roles in society and ultimately an impact on the nature of interpersonal relationships (Cook, 1990). When one's gender role is not well integrated a set of difficulties or type of maladjustment known as gender role conflict can result. For males, this kind of conflict becomes an issue when rigid, sexist, or restrictive gender roles lead to or promote personal restriction, devaluation, or violation of others or self (O'Neil, 1990). In the last decade, reseach by O'Neil and his colleagues on the Gender Role Conflict Scale (GRCS) have made it possible to measure four specific types of conflict. These gender role conflict patterns or factors consist of the following subscales: 1) Success, power, and competition issues, 2) Conflicts between work and family relations, 3) Restrictive emotionality, and 4) Restrictive affectionate behavior between men (O'Neil, Helms, Gable, David, & Wrightsman, 1986). Since different conflict patterns are expected to be associated with different personal and interpersonal problems, it seems reasonable that those with various subscale scores would request different types of assistance. Therefore, this study examines the relationship between the specific type of psychological help sought and the gender role conflict of clients as measured by the subscales of the GRCS. Subjects are male cadets at a Western Military Academy being seen in the campus counseling center. They have completed the Gender Role Conflict Scale upon in-processing for one of the following three types of counseling offered: 1) personal counseling, 2) alcohol counseling, and 3) leadership development counseling. We predict cadets with highest scores on the success, power, and competition subscale will be seen for leadership development counseling; those with highest scores on the restrictive emotionality and conflict between work and family relations subscales will be seen for personal counseling; and cadets with the highest scores on the restrictive affectionate behavior between men scale will be seen for alcohol counseling. The results and implications are discussed in terms of possibilities for addressing high risk populations, developing treatment programs, and guiding the counseling process to tap into issues that may not be readily disclosed.

Social Change and the Eating Habits Of Air Force Enlisted Personnel

Stephen J. Jirka, B.S.
St. Mary's University
Edna R. Fiedler, Ph.D.
Lackland AFB
Major Heather Ktenidis, R.D.
R. Brian Howe, M.A.
William G. Jackson, M.S.
Brooks AFB
Lt. Col. Diane Cortner, R.D.
Keesler AFB

Abstract

The stages of change that individuals progress through as they change habits have been applied to a variety of settings, including smoking cessation, weight management, and exercise. This study applied the stages of Prochaska's change model to the eating habits of 820 enlisted personnel eating habits before and during USAF basic training. One squadron's dining hall provided an "all healthy" diet (deleting high fat foods and increasing high fiber and calcium food choices) for 10 weeks (F=146, M=253), while another squadron provided the usual basic training fare (F=168, M= 253). Subjects in both groups completed the Health Habits and History Questionnaire, an inventory which translates actual food intake into nutrient intake, and a self assessment inventory of food intake based on Prochaska's model. Several areas of nutrient intake were measured: fat, cholesterol, fruit and vegetable, carbohydrate, protein, and calcium. Instruments were completed at the beginning and end of basic training. Results support Prochaska's model in that there were significant differences in actual nutrient intake by stage. Results also will be discussed in light of gender and group differences. However, even recruits who thought they were in the action and maintenance stages of Prochaska's model misjudged their actual food intake by as much as 80%. Results have strong implications for implementation of nutritionally sound strategies and policies as well as changing other life style habits.

Training Perishable Perceptual Skills

Steven J. Kass, Ph.D.
Robert H. Ahlers, Ph.D.
Naval Air Warfare Center
Training Systems Division

Abstract

As new training methods emerge (e.g., virtual reality, interactive courseware, multi-media), the desire to demonstrate their usefulness can tempt instructional developers to apply them inappropriately. It is of primary importance to understand the behavioral and cognitive aspects of a task and to design and implement the method that best satisfies the specific training requirement. For example, techniques appropriate for training cognitive tasks may not be effective for training predominantly perceptual tasks. One important, but difficult to train, perceptual task performed by U.S. Navy submarine officers is determining a contact's angle-on-bow as viewed through a periscope. This task is learned primarily on-the-job. Unfortunately, there is inadequate opportunity for on-the-job training, and there is a problem with determining the actual correct answer to provide as feedback. In order to meet this training need, the Naval Air Warfare Center Training Systems Division (NAWCTSD) developed a PC-based periscope simulator to provide the opportunity to practice perceptual tasks, such as determining angle-on-bow, contact identification and range estimation.

The purpose of the current study is to investigate the benefits of training perceptual tasks, such as periscope observation, using massed practice with immediate feedback. It is expected that individuals receiving practice with feedback will perform better than individuals receiving detailed task instruction and will perform as well as those receiving detailed instruction in addition to practice and feedback. Also, it is expected that these perishable perceptual skills will be retained longer when trained with repeated practice than when these skills are trained through instruction.

Summative Evaluation of OPS/CEAF FERL

Kurt C. Kraiger, Ph.D.
University of Colorado at Denver
Mark Teachout, Ph.D.
USAA
Theodore A. Lamb, Ph.D.
USAF Armstrong Lab USAFA/OL

Abstract

In order to support the development and implementation of the OPS/CEAF FERL program, the assessment team developed a multi-faceted evaluation approach. The summative evaluation was conducted primarily by researchers based at the University of Colorado at Denver, with the assistance and cooperation of both Civil Engineering faculty at the Air Force Academy and researchers at Armstrong Laboratories. The objectives of the summative evaluation were to determine whether learning occurred during OPS/CEAF FERL, to determine whether this learning transferred to cadet performance during subsequent courses during their junior and senior years, and to determine if this learning will transfer to job performance during their first assignments. To accomplish these objectives, a number of tests and measures were developed. These measures were based on a typology of learning outcomes presented in Kraiger, Ford, and Salas (1993), and included: a FERL-Knowledge Test (assessing specific knowledge drawn from the learning objectives for OPS/CEAF and each FERL activity); an assessment of cadets motivation to learn, a test of cadets confidence (that they could perform activities demonstrated during FERL when on the job), a step-analysis test (assessing cadets deeper understanding of procedural knowledge related to CE activities), and an assessment of cadets structural knowledge (or knowledge organization) for key concepts. These tests were administered before and after OPS/CEAF FERL in both 1994 and 1995. Additionally, all tests were administered at the end of subsequent courses during cadets junior or senior years. Grades in courses and instructor ratings of student attitudes in these courses were also tracked.

Key results showed that learning occurred during the OPS/CEAF FERL in both years. Most tests showed significantly higher scores at post-test than pre-test. The percent increase in scores on specific tests was often over 100%. Cadets completing the OPS/CEAF FERL program also had higher pre-test scores (prior to subsequent courses) on the assessment tests than did environmental engineering students who had not completed the course. The results for transfer of training to the classroom were mixed, but showed some evidence that cadets motivation, confidence, and knowledge for procedures were related to classroom attitudes and achievement.

References

Kraiger, K., Ford, J.K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation

[Monograph]. Journal of Applied Psychology, 78, 311-328.

Application of the Critical Incident Technique to Enhance Crew Resource Management Training

Kristi K. Logan
Mary Ann Hanson
Jerry W. Hedge
Kenneth T. Bruskiewicz
Walter C. Borman
Personnel Decisions Research Institutes, Inc.

Frederick M. Siem
Armstrong Laboratory Human Resources Directorate

Abstract

This research project employs the "critical incident" technique to build a foundation for developing improved crew resource management (CRM) training tools. In the first phase of the project, this critical incident technique was used to collect a large number of "performance examples" from Air National Guard (ANG) tanker units. These performance examples are descriptions of effective, ineffective, and average levels of CRM performance. Some of these performance examples are descriptions of things individuals have done and some describe the performance of teams, where teams are defined as two or more individuals working together. Once a large number of performance examples had been collected, additional aircrew members were asked to assign each performance example to a dimension of CRM-related performance and rate the effectiveness of the behaviors described. These "retranslated" performance examples provide the building blocks for developing behavior-based rating scales targeting CRM dimensions of crew and individual performance. These scales can be used to obtain accurate assessments of CRM performance before, during, and/or after CRM training. Current activities include completion of rating scale development for all CRM dimensions of team performance as well as for each individual crew position. A preliminary training needs analysis has already been conducted, and future plans include additional needs analysis work to help guide the development of CRM training tools. Basic and field research studies will also be conducted to develop a better understanding of CRM performance and the relationship between individual and crew performance. All of this information will provide a foundation for the new CRM training tools to be developed for Air Force tanker crews.

Perceived Job Security and Veteran Status: Effects of Type of Occupation

Michael D. Matthews, Ph.D.
Drury College
Charles N. Weaver, Ph.D.
St. Mary's University

Abstract

An analysis of national survey data showed that veterans of military service were more likely than non veterans to perceive difficulty in locating a comparable job if forced to do so. Further analyses revealed that this difference was found for both white and blue collar workers. Moreover, type of occupation was also related to responses, with veterans employed in managerial positions, clerks, and the service industry reporting more difficulty in finding a comparable job than non veterans, and no difference with respect to this variable among professional workers, sales personnel, craftsmen, equipment operators, laborers, and farmers. Implications are discussed.

Matthews and Weaver (1992) reported the results of a preliminary study which showed that while veterans of service in the United States military do not perceive a greater risk of losing their civilian job than non veterans, they do perceive greater difficulty in locating a comparable job if forced to do so. This perception may have very real consequences in the likelihood of the employee seeking another job, or with job satisfaction if the availability of other options is perceived to be minimal.

A limitation of Matthews and Weaver's (1992) study was that it did not control for occupational and demographic variables that might interact with veteran status to affect job attitudes. It is known, for example, that type of occupation has a strong relationship to what qualities a person expects in a job (e.g., Weaver & Matthews, 1987). Other variables that may impact job attitudes include age, ethnicity, sex, and educational levels, to name a few (see, for example, Muchinsky, 1987). It would be useful to examine the relationship between perceptions of job security and veteran status as a function of such variables.

The purpose of the current study was, therefore, to control for demographic variables in order to clarify the relationship between veteran status and perceived job security. More specifically, the present study compared the responses of veterans and non veterans as a function of type of occupation and occupational class (white collar versus blue collar). The results should be of value to managers and other personnel specialists who employ veterans.

Using Pre-course Reflective Judgment and Critical Thinking Measurement Instruments as Guides for Lesson Preparation.

Ronald F. K. Merryman, M.S.
Anthony J. Aretz, Ph.D.
United States Air Force Academy

Abstract

Often, teaching methods which encourage growth of critical thinking abilities are employed in the college classroom. However, applying these generically to all students in every section without full appreciation for their actual level of reflective judgment could be analogous to attempting to teach a car mechanic how to change a tire; or expecting a novice reader to expound on the intricacies of Shakespearean conflict. In both instances, the dialogue is not at the level of the student. Therefore, assessing the students' level of reflective judgment prior to the course is an important step in optimizing their strengths and developing their weaknesses. Instructors can use this data to develop specific questions for course objectives, lessons, sections and essays.

This study involved administering a paper and pencil reflective judgment instrument (RJI) to United States Air Force cadets who volunteered to participate in the study. The subject pool consisted of 50 freshmen cadets at the US Air Force Academy currently enrolled in "Introduction to Psychology" in the Department of Behavioral Sciences and Leadership. The RJI required twenty minutes to complete and was analyzed using Excel 5.0 on a Intel 80486 processor-based computer. Results of the RJI will be used to (1) determine the collective pre-course reflective judgment level for each of three sections of students; and (2) develop and modify lesson plans to capitalize on opportunities for discussion and critical thought about course material.

References

Diefenderfer, K.K. (1993). Critical thinking skills: Recent research and classroom applications. The National Institute on the Teaching of Psychology, 15.

King, P.M. (1992, January/February). How do we know? Why do we believe? Learning to make reflective judgements. Liberal Education, 78, 1. 2-9.

Lynch, C.L., Kitchener, K.S., King, P.M. (1995). Developing reflective judgment in the classroom. New Concord, KY: Reflective Judgment Associates.

Characterization of Sleep, Mood, and Performance Patterns in Battalion Staff Members at the Joint Readiness Training Center

Robert J. Pleban, Tina L. Mason, and Patrick J. Valentine
U.S. Army Research Institute for the Behavioral and Social Sciences
Fort Benning, Georgia

Abstract

The battlefields of the future are likely to be chaotic, intense, and extremely lethal. New technologies combined with the adoption of a force projection doctrine have greatly increased the tempo of modern warfare. Advances in target acquisition, communications, sensor systems, and the almost complete mechanization of land combat forces enable units to fight longer, harder, and faster than ever before. Continuous operations, the type of high intensity, fast paced operations described above, will put tremendous stress on the soldier's recuperative abilities. Opportunities for sleep, while possible, will be brief or fragmented in such situations. Individuals at most risk during continuous operations are those with heavy cognitive work load requirements. Soldiers who must process and evaluate large amounts of information such as fire direction center crews, radar operators, staff tactical operation center members, and leaders at all levels are most susceptible to the effects of continuous operations.

The objective of this research was to assess the sleep-work schedules of an actual battalion staff at the Joint Readiness Training Center (JRTC). Specific focus was directed toward staff members' sleep habits both during and off-rotation; identification of performance and mood changes during the course of the rotation; and evaluation of the effectiveness of current staff members' sleep schedules in sustaining performance for extended periods of time.

Ten U.S. Army battalion staff members were outfitted with wrist-worn, solid state activity monitors and tracked across a 16-day exercise. Sleep-work patterns were assessed along with performance on a synthetic work battery, and mood state. Records from the activity monitors indicated that the average daily sleep obtained by the staff members was 5.2 hours (range 3.5 - 6.4 hours). The staff averaged almost three hours less sleep per day than was needed for total recovery. Certain staff positions received very little sleep across the exercise. Over sixty percent of the sleep obtained was fragmented in nature (sleep periods of 10 minutes or less). With regard to overall performance, no significant negative changes in individual performance were observed over time. However, substantial increases in response variability were noted for one staff member. Subjective reports indicated that sixty percent of the staff members felt that their work load was excessive.

Better distribution of duties among staff members and other unit personnel were viewed as critical for sustaining effective staff performance for extended periods. In addition, commanders must take an active role in the development of unit sleep/work management plans. This includes educating unit members on the importance of sleep in combat operations and how to optimize the recuperative value of available sleep periods through specifically tailored unit sleep plans.

Scoring Work Sample Data of Complex Ill-Structured Tasks with Expert Holistic Judgments

Robert A. Pokorny, Ph.D.
Air Force Armstrong Laboratory

Abstract

Many situations require behavioral scientists and supervisors to measure a worker's ability to perform cognitively-demanding ill-structured tasks, such as troubleshooting. To measure can-do performance (as opposed to will-do performance), work sample tests provide more direct information than other commonly used performance measures, such as supervisor or peer ratings, or job knowledge tests. But work sample tests are difficult to score. Borman (1991) states that the scoring system of work samples must be understood unambiguously at an operational level--work samples which are difficult-to-score, he suggests, should be avoided. In a variety of situations, however, such as evaluating the effectiveness of a training system (Boyer, Hall, Rowe, & Pokorny, 1996), scientists and practitioners are faced with the need to measure performance of difficult tasks, with little time and money to complete the evaluation. To solve this problem, this poster describes how to measure work samples of difficult tasks cheaply and quickly. The crux of the solution is to have expert raters, already familiar with the work sample task to be rated, read records of the work samples and assign holistic ratings which reflect the quality of each work sample. Despite some researchers' possible misgivings about experts' holistic ratings, these ratings can provide reliable performance measurement: in this study, the ratings of different experts correlated in the .90s. This poster describes how experts can provide good holistic ratings of complex ill-structured tasks.

References

Borman, W. C., (1991). Job Behavior, Performance, and Effectiveness: In M. D. Dunnette & L. M. Hough (Eds.), Handbook of industrial and organizational psychology: Vol. 2. (2nd ed., pp. 271-326). Palo Alto, CA: Consulting Psychologists Press.

Boyer, B. S., Hall, E. M., Rowe, A. L., & Pokorny, R. A., (1996, April). No Pain, No Gain: The Effect of an Intelligent Tutoring System on F-15 Troubleshooting Performance. Paper presented at 15th Applied Behavioral Science Symposium, Colorado Spring, CO.

Implementation of Assessment Measures and Curriculum Integration

Capt Michael P. Rits, USAF
HQ USAFA/DFCE
Mark Teachout, Ph.D.
USAA
Theodore A. Lamb, Ph.D.
USAF Armstrong Lab USAFA/OL

Abstract

The Department of Civil Engineering recognized the need for improving the curriculum, particularly through formal curriculum integration, and began developing an entry-level civil engineering course, CE 351, which it hoped would help accomplish that task. An agency outside of the Academy was sought to objectively conduct a formal assessment of the new course and the resulting changes to the curriculum. In January 1994, the Armstrong Laboratory agreed to conduct this multi-year evaluation (1994-1998), bringing on board nationally recognized experts in the field of instructional technology and industrial psychology for both formative and summative assessment. The goal of the formative assessment was to evaluate and improve CE 351, while the summative assessment was designed to measure the improvements in the students' performance (knowledge, skills and attitudes), both during the new course and throughout the curriculum.

The entire department has been committed to this assessment effort, and participates in two department assessment/integration workshops each year, facilitated by the assessment team. During the Fall workshops, the department focus is on implementing changes to CE 351 based upon the formative assessment data analysis. The focus of the Spring workshop is curriculum integration and improvement, predominantly based upon discussions stemming from the summative assessment data analysis. All the assessment instruments have been developed by working directly with the faculty and when changes are made to the curriculum, the assessment instruments are modified accordingly. The instructors include class time in their syllabi for administration of assessment instruments and are committed to integrating the experiences from CE 351 into their courses. While not all summative data has been collected in the assessment process, the faculty have already begun to see definite improvement in the learning demonstrated by cadets having gone through the new curriculum.

Short Term Effects of Acceleration on Human Subjects

Lieutenant Dylan Schmorrow, Ph.D.
James Siwert
David Moyers
Naval Air Warfare Center - Aircraft Division

Abstract

This study examined the short term effects of acceleration exposure on human test subjects, involving 1,333 subject centrifuge exposures in the Dynamic Flight Simulator at the Naval Air Warfare Center - Aircraft Division Warminster, PA. These exposures covered a period of 7 years and 17 different acceleration research projects. Acceleration levels ranged from -1 Gz to +12 Gz, with the average level of 8.5 +Gz. The human test subject pool for this period consisted of 84 males and 8 females. Following each complete set of runs, subjects were asked if they experienced fatigue, nausea, disorientation, headache, pain, and muscle strain. It was also noted if they experienced a G induced loss of consciousness episode. From this data a historical analysis was done comparing these short term effects based on different types of G exposure and G levels. Other variables considered for this analysis were length of exposure and age. Percentage and overall numbers of each variable experienced were documented and compared. In addition, the average length of exposure time for each individual project along with a brief description of the projects are included.

An Enlistment Screen for Non High School Graduates: Development, Operational Test, and Evaluation[8]

Thomas Trent
Navy Personnel Research and Development Center

High school dropouts comprise a large portion of potential recruits for military enlistment. Yet, the Services minimize the numbers of nongraduates in order to control enlistment attrition and to enhance the quality characteristics of recruits. The basis for this policy is also found in the civilian literature: in addition to low educational achievement, nongraduates experience more criminal involvement, drug and alcohol abuse, unemployment, and psychological problems. Despite an enlistment attrition rate of over 50%, the Navy Recruiting Command accepts limited numbers of nongraduates who score at or above the 50th percentile on the Armed Forces Qualification Test (AFQT). Thus, the objective of the research was to design a recruiter-administered model of individual differences for these "B-cell" applicants.

The model development sample included 25,199 Navy enlisted personnel who had not completed a traditional high school diploma. The evaluation sample consisted of 3,217 accessions who qualified on the Compensatory Screening Model (CSM) and a control group of 1,086 nongraduates who were exempted. The criterion measure was completion of the first two years of enlistment versus premature separation. Parameter estimates were derived from a logistic regression of service completion on variables available from the Military Entrance Processing Reporting System (MEPRS).

The CSM was operationalized as an actuarial table of two-year service completion probabilities. The screening variables were years of education, type of secondary education credential, age, and AFQT score. As compared to the control group, attrition during the first 18 months of enlistment was found to be 6 percentage points lower for the CSM-screened group. Most of this difference was due to fewer drug-related and misconduct discharges. Compared to FY88-91 nongraduates, the CSM-screened group also completed more years of education and alternative secondary credentials. The majority scored in the two highest AFQT categories.

Variance within attrition-related variables make additional nongraduate applicant screening feasible. The Navy will recruit at least 2,800 B-cells in FY96. At that level, attrition reduction from CSM screening is expected to save $3.4M in training costs. Furthermore, the enlistment of high aptitude nongraduates will enable recruiters to fill undermanned technical occupations.

---

[8] The opinions expressed in this abstract are those of the author, are not official, and do not necessarily reflect the views of the Navy Department. The author is grateful to Steven Devlin for his contributions to this project.

Application of Sequential Data Analysis Techniques to the Instructional Domain

Brenda M. Wenzel
Mei Technology
San Antonio, Texas

Data reduction techniques for analyzing sequential data provide a novel approach to describing, comparing, and evaluating instructional media. This paper describes how the approach was used to compare traditional Air Force classroom instruction to computer-based instruction (CBI) developed with the Experimental Advanced Instructional Design Advisor (XAIDA). XAIDA is an authoring system for developing computer-based maintenance training. XAIDA is designed to capture a subject matter expert's domain-specific knowledge about a device or system and use it to automatically generate CBI lessons on the maintenance of that device or system. XAIDA is being developed and tested under the sponsorship of the AF Armstrong Laboratory as part of a research program on automating instructional design.

Available to us as data are video tape transcripts, computer journals that automatically capture human-computer interactions, and direct observation. The steps to conducting sequential data analysis, once the data has been collected, involve: (1) identifying and defining observable categories that are of practical or theoretical interest; (2) coding the data into the categories, (3) building a data matrix of the frequency of co-occurrence between categories, (4) calculating observed transition probabilities for all cells in the matrix, (5) graphically representing the transition matrix, and (6) conducting statistical tests within and between transition matrices. Sequential data analysis techniques provide a means of examining the manner and degree that instruction is sequenced rather than random, identifying the differences in instructional sequences between instructional media, and exploring the impact that instructional sequence has on learning effectiveness.

-A-
Ahlers, R. H.: 374
Albright, R. R.; 207
Allen, R. C.; 161
Andre, A. D. 105
Aretz, A. J.; 149, 123, 378
Asiu, B.; 64
Austin, J. A.; 362

-B-
Bailey, L.L.; 334
Baker, H. G.; 256
Baker, J. S.; 367
Barth, T. S.; 362
Bauer, D. H.; 143
Bennett, T.; 284
Bennett, W. R., JR.; 103, 268, 363
Benson, M. J.; 44
Berger, R. C.; 70
Beringer, D. B.; 130
Blankman-Alexander, D. M.; 362
Borman, W. C.; 364, 376
Boyce, L.; 82
Boyer, B. S.; 76
Bruskiewicz, K. T.; 364, 376

-C-
Cacioppo, A. J.; 137
Carr, W. J. ; 353
Chaiken, S. R.; 179
Cleveland, J. N.; 227
Cornum, K. G.; 143
Cortner, D.; 373
Craiger, J. P.; 305
Croxton, C. A.; 70
Cummings, W. H., III; 365
Curphy, G. J.; 206, 219
Curry, G.; 284

-D-
Dansby, M. R.; 310, 317, 322
Dobie, T. G.; 110
Dutcher, J. S.; 305

-E-
Eckerman, J. R.; 372
Elliot, G. S.; 165

-F-
Fiedler, E. R.; 373
Fore, T. A.; 366
Frieman, S. R.; 235

-G-
Garris-Reif, R.; 368
Garvin, J. D.; 340
Gibb, G.D; 346, 367
Ginnett, R. C.; 362
Gugerty, L.; 179
Gustafson, K. L.; 103, 363

-H-
Hagman, J. D.; 170
Hah, S.; 97
Hall, E. P.; 76
Hampton, S.; 346
Hansen, J. A.; 52
Hanson, M. A.; 346, 376
Hayes, R. T.; 368
Hedge, J. W.; 364, 376
Henderleiter, H. M.; 289
Hendrix, W. H.; 369
Henle, C.; 227
Herd, A. M.; 82, 240, 370
Holt, D. T.; 261
Hosmer, C.; 28, 87
Houlihan, C.; 193
Howe, R. B.; 373
Hoyle, M.; 317
Hughes, H. M.; 293
Hughes, R. L.; 369
Huhn, B. P.; 118
Hurry, L.S.; 181
Hurwitz, J. B.; 179

-I-
Irwin, C. M.; 362, 371

-J-
Jackson, R. J.; 372
Jackson, W. G.; 373
Jandzinski, G.; 246
Jirka, S. J.; 373
Johannsen, C. T.; 149
Jones, G. V.; 11

-K-
Kanki, B.: G.; 362
Kass, S. J.: 374
Kelly, P. T.; 207
Kelly, S. C.; 367
Kern, T.; 340
Kieling, C. R.; 372
King, R. E.; 1
Knouse, S. B.; 322

Kopania, T. P.; 32
Kraiger, K. C.; 103, 375
Ktenidis, H.; 373
Kyllonen, P. C.; 179

-L-
Landis, D.; 317
Lamb, T. A.; 103, 375, 381
Landry, J. R.; 187
Law, C.; 82
Lawson, K. L.; 261
Lindsley, D. H.; 23, 298, 353
Lofgren, S. T.; 261
Logan, K. K.; 364, 376

-M-
Marble, D.; 216
Martin, M.; 6, 11
Mason, T. L.; 379
Massey, M.R.: 251
Matthews, M. D.; 377
May, J. G.; 110
McCormick, E. P.; 156
McCown, D. L.; 91
McGlohn, S. E.; 1
McIntire, F.; 193
McKay, D.; 359, 361
Merryman, R. F. K.; 137, 378
Micalizzi, J.; 200
Mills, V. K.; 206
Mitchell, J. L.; 273, 279
Morgan, K.; 227
Moyers, D.; 382
Murphy, K. R.; 227

-N-
Nappi, M. J.; 362
Nason, E. R.; 362
Niebuhr, K. E.; 322
Niebuhr, R. E.; 322

-O-
Ober, K. R.; 123
Orth, M.; 227
Osten, K. D.; 219

-P-
Packard, G. A.; 70
Perrin, B.; 268
Phalen, W. J.; 273
Pittman, T. S.; 211
Pleban, R. J.; 379
Pokorny, R. A.; 76, 380
Pode, P. E.; 143
Poole, P.E.: 143
Porter, D. B.; 44, 52, 58
Porter, R. D.; 356
Portman, C.; 361
Potter, E. H., III; 211
Preston, K.; 58
Pringle, H.; 370
Prochaska, F.; 213

-R-
Reinsma, J.; 217
Rezlaff, P. D.; 1
Rits, M. P.; 103, 381
Rowe, A. L.; 76
Rosenbach, W. E.; 211
Rosengren, B.; 218
Rue, R. C.; 279
Rueb, J. D.; 91

-S-
Schmorrow, D.; 359, 361, 382
Sellman, W. S.; 353
Shane, G. S.; 181, 251, 261
Shappell, S.; 360
Shaw, R. V.; 334
Siem, F. M.; 364, 376
Simon, X.; 367
Simpkins, P. S.; 362
Singer, M. J.; 161
Siwert, J.; 382
Sola, D.; 367
Stanny, R.; 357
Sterling, B.; 175
Sterling, J. L.; 356
Stiles, R.; 82
Stokan, L. A.; 97
Stone, B. M.; 279, 284
Strunk, V. L.; 214

-T-
Teachout, M.; 375, 381
Thompson, B. R.; 328
Thoreson, S. A.; 279
Trent, T.; 383
Tu, D. S.; 105
Turner, K. L.; 279, 284

-V-
Valentine, P. J.; 379
VanScotter, J. R.; 181, 246, 251
Vaughan, D.; 246
Voetberg, J.; 219, 220

Vroonland, J. P.; 356

-W-

Wabiszewski, C.; 367
Waggle, M. V.; 17, 38
Wallace, S. G.; 356
Wallisch, B.; 215
Walter, K. E.; 317
Waters, B. K.; 298
Weaver, C. N.; 377
Wenzel, B. M.; 384
Wheeler, W.; 363
Wickens, C. D.; 156
Wiegman, D. A.; 358
Wilcove, G. L.; 305
Wise, J. A.; 346
Wiskoff, M. F.; 23
Wolf, J.; 346

-XYZ-

Yadrick, R. M.; 268