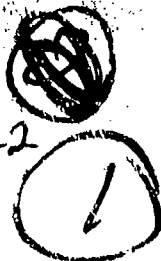


AD-A253 006



US AFA-TR-92-2

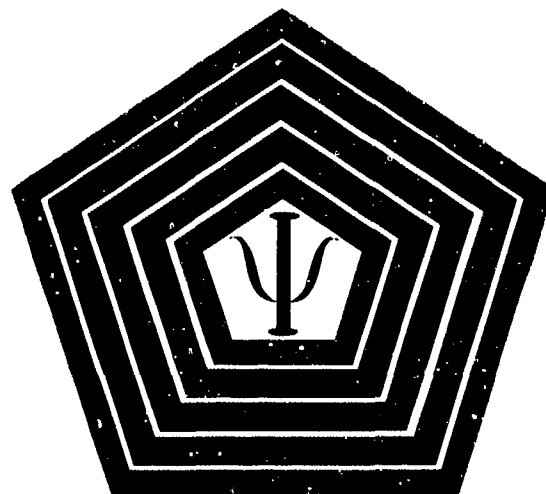


Proceedings

PSYCHOLOGY

in the

DEPARTMENT OF DEFENSE



DTIC
S ELECTE D
MAY 12 1992

THIRTEENTH SYMPOSIUM

15-17 APRIL 1992

This document has been approved
for public release and sale; its
distribution is unlimited.

Sponsor

USAF Academy Department of Behavioral
Sciences and Leadership

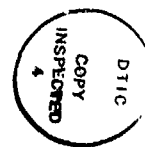
92-12243



92 5 05 019

USAFA-TR-92-2

PROCEEDINGS OF THE THIRTEENTH SYMPOSIUM
PSYCHOLOGY IN THE DEPARTMENT OF DEFENSE
15-17 APRIL 1992



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

Accession For	
NTIS CRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and/or Special
A-1	

SYMPOSIUM CHAIR
LIEUTENANT COLONEL FREDERICK W. GIBSON



**PROFESSOR AND HEAD, DEPARTMENT OF
BEHAVIORAL SCIENCES AND LEADERSHIP**

Colonel Richard L. Hughes

EDITOR

Captain Anthony J. Aretz

PROGRAM COMMITTEE

Major William T. Wittman, Chair
Lieutenant Colonel Benjamin L. Dilla
Major Gail Hinesley
Captain Jay A. Horn
Captain Harold S. Hillman

OPERATIONS

Major George R. Mastroianni, Chair
Captain Kevin A. Ashley
SSgt Debra A. Barren
Ms Diane Nicoletta

PUBLICITY

Major R. Jeff Jackson

PROTOCOL

Captain John M. Barker Jr.

REGISTRATION

Captain Christopher T. Sorrentino
Captain Dean H. Orrell II

LODGING

Captain Cathy C. Clothier

TRANSPORTATION

Captain Ardis Cecil

FINANCE

Captain Jay A. Horn

DATABASE

Captain Andrew G. Stricker
Captain Richard Tallarigo

REPORT DOCUMENTATION PAGE

Form Approved
 OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED			1b. RESTRICTIVE MARKINGS		
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for Public Release and Distribution Unlimited.		
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE			4. PERFORMING ORGANIZATION REPORT NUMBER(S) USAF TR 92-2		
6a. NAME OF PERFORMING ORGANIZATION Department of Behavioral Sciences and Leadership			6b. OFFICE SYMBOL (If applicable) DFBL		7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) U.S. Air Force Academy Colorado Springs, CO 80840-5701			7b. ADDRESS (City, State, and ZIP Code)		
8a. NAME OF FUNDING/SPONSORING ORGANIZATION		8b. OFFICE SYMBOL (If applicable)		9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	
8c. ADDRESS (City, State, and ZIP Code)		10. SOURCE OF FUNDING NUMBERS			
		PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
11. TITLE (Include Security Classification) Proceedings, Psychology in the Department of Defense, Thirteenth Symposium (UNCLASSIFIED)					
12. PERSONAL AUTHOR(S) ANTHONY J. ARETZ, Capt, USAF					
13a. TYPE OF REPORT Proceedings		13b. TIME COVERED FROM _____ TO _____		14. DATE OF REPORT (Year, Month, Day) 1992, April, 15	
15. PAGE COUNT 306					
16. SUPPLEMENTARY NOTATION					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)		
FIELD	GROUP	SUB-GROUP	Symposium, Training, Leadership, Counseling, Cognitive Performance, Family Issues, Total Quality Management, Desert Storm, Human Factors, Gender Issues		
05	05				
05	10				
19. ABSTRACT (Continue on reverse if necessary and identify by block number) These proceedings include papers and presentations that deal with a wide range of research in psychology with emphasis on military issues.					
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED		
22a. NAME OF RESPONSIBLE INDIVIDUAL ANTHONY J. ARETZ, Capt, USAF			22b. TELEPHONE (Include Area Code) (719) 472-3860		22c. OFFICE SYMBOL DFBL

USAFA-TR-92-2

**Technical Review by Lt Col Tom McCloy
Department of Behavioral Sciences and Leadership
USAF Academy, CO 80840**

**Technical Review by Maj Steve Green
Department of Management
USAF academy, CO 80840**

**Editorial Review by Lt Col Donald Anderson
Department of English
USAF Academy, CO 80840**

This research report entitled "Proceedings of the Thirteenth Symposium Psychology in the Department of Defense" is presented as a competent treatment of the subject, worthy of publication. The United States Air Force Academy vouches for the quality of the research, without necessarily endorsing the opinions and conclusions of the authors.

This report has been cleared for open publication and public release by the appropriate Office of Information in accordance with AFR 190-1, AFR 12-30, and AFR 80-3. This report may have unlimited distribution.

Robert K. Morrow, Jr.
**ROBERT K. MORROW JR., Lt Col, USAF
Director of Research**

17 MAR 92
Dated

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 unlimited, in accordance with IVA, AFR 80-45.

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

Acknowledgements

Cover

Graphics Division of Audiovisual Services
Dean of Faculty, USAF Academy

Administrative Support

Sheila Rhett
Terri Newcomb
Dawn Anthony

**PROCEEDINGS OF THE THIRTEENTH SYMPOSIUM ON PSYCHOLOGY
IN THE DEPARTMENT OF DEFENSE**

Page

PAPER PRESENTATIONS

FLIGHT SICKNESS

Multi-Modal Assessment and Treatment of Airsickness R. Jeffrey Jackson	1
Airsickness Management and the Benefits of Gyro-Fitness Training William T. Wittman Carl J. Mallery Robert C. Berger George R. Mastroianni Robert J. Schoen	6
Flight Simulator-Induced Sickness and Visual Systems Evaluation Halold D. Warner Gary L. Serfoss Todd M. Baruch David C. Hubbard	11

TQM APPLICATIONS

Applying Total Quality to a Statistics and Research Design Course: A Case in Points John Micalizzi David B. Porter	16
A Soft-Systems Model for Continuous Improvement Andrew G. Stricker Thomas McCloy	21
Student Course Critiques: A Case Study in Total Quality Education David B. Porter	26

CADET LEADERSHIP

Does the United States Air Force Academy's Training Philosophy Really Work? Michael A. Rosebush	31
Cadets' Perceptions of Leadership Development Activities at the United States Air Force Academy Derek A. West Benjamin L. Dilla	36

GENERAL TQM

The Quintessence of Quality	
David B. Porter	41
Systems Science in Support of Total Quality Management	
Larry B. Rainey	
Charles J. Yoos	46
Total Quality Management: A Unique Challenge for Behavioral Science	
Researchers	
Jimmy L. Mitchell	
J. K. Knight	51

COOPERATION AND COMPETITION

Student Attitudes Toward Goal Interdependence Characterizations of	
Competition and Cooperation	
R. Joe Grisham	56
Gender and Attitudes as Predictors of Video Game Performance and	
Satisfaction	
Brittany J. Thurber	
David B. Porter	
Anne C. Parker	61

PERSONNEL TESTING AND SELECTION

Operational Implementation of a Validated Personnel Selection System for	
Landing Craft Air Cushion Vehicle Operators	
Daniel L. Dolgin	
David R. Street	
Tatree Nontasak	
David J. Blower	
Kathleen Travis	66
The Role of Personnel Testing in a Changing Air Force	
Constance L. Riggs	
Tina S. Ziebell	71
Improving Career Selection with the Armed Services Vocational Aptitude	
Battery 18/19 Career Exploration Program	
Kenneth W. Wegner	76

LEADERSHIP

Military vs. Business Leadership: Suggestions from Career Description Analysis	
David D. Van Fleet	
Tim O. Peterson	81
An Exploration of Transformational Leadership and Its Effects	
Ruediger Mueller	
Marshall Sashkin	
William E. Rosenbach	86

ORGANIZATIONAL ISSUES

Centrality of Work: The General Moderator Hypothesis and Scale Development	
Frederick W. Gibson	
Kevin A. Ashley	
Douglas R. Lindsay	91
Do Situational Variables Affect Adherence to Formal Organizational Ethics Policies?	
Gordon J. Curphy	
Frederick W. Gibson	
Gary Macomber	
Callie Calhoun	96
Perceived Job Security as a Function of Prior Military Service	
Michael D. Matthews	
Charles N. Weaver	100

PERSONNEL MEASUREMENT

Age Group Differences and Estimated Frequencies of the MBTI Types: Proposed Changes (M I N INTJ Forever?)	
William H. Cummings	104
The Myers-Briggs Type Indicator and Acquisition Management	
Carl Bryant	109

DISPLAY ISSUES

Visual Representation Study	
Blair W. Conroy	
Jay A. Horn	
Paul D. Grunzke	
Christopher D. Wickens	113

	PAGE
Distance in Multifunction Displays Karen Seidler Christopher D. Wickens	118
Visual Field Versus Muscle Synergy Effects on Control/Display Compatibility: Seeing is Believing Dennis B. Beringer Charles J. Worringham	123
<u>AIRCREW ISSUES</u>	
Aircrew Experiences During the Vietnam Conflict: Combat, Cohesion, and Leadership John A. Ballard Antone E. Gajeski	128
Aircrew Coordination: What Does it Take? Rene'e Stout Carolyn Prince David P. Baker Maureen L. Bergondy Eduardo Salas	133
<u>TRAINING AND PERFORMANCE THEORY I</u>	
Attention and Skill Anthony J. Aretz	138
Virtual Reality and Education Christopher D. Wickens	143
<u>TRAINING AND PERFORMANCE THEORY II</u>	
Theories of Expertise as Models for Understanding Situation Awareness Peter M. Crane	148
A Structural Measure of Learning for Ill-Structured Domains John Landry	153
Performance Decrements During Sustained Operations: The Role of practice and Task Structure D. G. Humphrey A.F. Kramer R. R. Stanny K. Travis	158

PERFORMANCE IMPROVEMENT**Ergonomic Computer Work Environment**

Brian Keller	
Todd McDowell	
Kelly Kirkpatrick	
Tom Becker	
Jay Horn	
Dean Orrell	
Mike Strevell	163

Development of an Instructional Tutor for the T-41 Flight Program

Steven A. Peeples	
Jay A. Horn	
William A. Jones Jr.	168

Improving Visual Acuity Performance

Jennifer L. Hughes	
Leray L. Leber	
Amy L. Walker	173

GENERAL EXPERIMENTAL**Effects of Image Update Rate on Target Identification Range**

Elizabeth L. Martin	
Julie Mapes Lindholm	178

Individual Differences in a Dual Task Context

Christopher R. Rate	
Anthony J. Aretz	183

JUDGEMENT AND DECISION MAKING**Illusions of Unique Invulnerability: Implications for Selection of Military Personnel for Risky or Hazardous Duties**

Kristen Morton Vance	188

Inertia and Adaptation in the Sequential Processing of Information

Elliot E. Entin	193

Future Research Applications for a Job Aiding/Training Conceptual**Methodology**

Kara B. Worthington	
Theodore A. Lamb	198

GENDER ISSUES I

Gender Differences: Stress Effects on General Well-Being and Absenteeism	
William H. Hendrix	
Gail S. Gibson	202
Determinants of Selective Processing of Emotion Cues	
Maryanne Martin	
Gregory V. Jones	207

GENDER ISSUES II

Women in the Military: Comparison of Attitudes and Knowledge of Service	
Academy Cadets Versus Private College Students	
Michael D. Matthews	212
Gender and Performance in Naval Aviation Training	
Annette G. Baisden	217

DESERT SHEILD/STORM

Mental Health Issues: Lessons Learned from Operation Desert Sheild/Storm	
Jame A. Martin	221

POSTER PRESENTATIONS

Organizational Readiness for TQM	
Jerry A. Arends	226
Racial Differences in Naval Aviation Training	
Annette G. Baisden	227
Work Attitudes Six and Twelve Months Following Implementation of a	
Restrictive Smoking Policy at a Military Hospital	
John A. Ballard	
Kenneth E. Olive	
Guy S. Shane	
Mark E. Kain	228
Validity of the Structural Scales of the CPI for Predicting the Performance of	
Junior Officers in the U. S. Coast Guard	
Rex J. Blake	
Earl H. Potter III	
Richard E. Slimack	229

	PAGE
The Effects of Uniform Type and Gender on Helping Behavior Tammy Carlgren Rick Cooney Mike Pine	230
Prediction of Absenteeism, Turnover Intention, Job Satisfaction and Performance Using Two Motivational Models: Goal Setting and Job Design Terri L. Coccia	231
Developing Learning Skills: The Effectiveness of a Videotaped Course Richard Cuervo	232
Pressure Suit Design and Extra Vehicular Activity Crista Davis	233
Hemispheric Lateralization in Left Handed People Danile C. Engberson	234
Effects of Contextual Cue Degradation and Experience Level on Performance Dayne R. Fentress Janelle K. Viera	235
Analysis of Requisite Abilities of Air Traffic Control Specialists: Toward an Effective Screening System. Gerald D. Gibb Marvin L. Smith Neil Swindells Dieter U. Haney	236
Cigarette Smoking During Operation Desert Storm Robert K. Gifford Richard T. Keller Ernest Degenhardt	237
Teams - A Technique of Formation Jay W. Gould III	238
External Validity in Performance Measurement Research F. Douglas Hoicombs	239
Navy Recruit Psychological Screening Imelda Idar James A. Scarnozzino	240
Schooling and Training Approaches Measurement Project (STAMP) William J. Jones	241

Learning Skills Center Survey Tiffany N. Jordan Valerie J. Matthews	242
History of Airpower Videodisc Program Evaluation Jerry Lautenschlager Patrick Patterson Dean Orrell	243
The Effect of Feedback Content on Group Practice Behavior and Individual Achievement Victoria Leitham Michael M. White	244
Assessing Test-Retest and Inter-Rater Reliabilities of the Worker Rehabilitation Questionnaire (WRQ). David D. Robinson Kristie Prue John Howes	245
Comparisons of Combat Stress Reactions in Two Soldiers Populations During Hostilities in the Gulf War Robert R. Roland	246
Causal Analysis of Leader Behaviors Richard S. Tallarigo	247
Factor Structure Comparisons of Behavioral and Trait Assessments of Leadership Richard S. Tallarigo	248
Total Quality Management and the Training Philosophy in Cadet Squadron Eleven Sterling E. Tree Steven A. Higgins Jay A. Horn Jeffery S. Austin	249

PANEL PRESENTATIONS

Sexual Harassment in the Military

Robert H. Faley
Dan Landis
Robert E. Nelbuhr
Sharon L. Oswald
Mickey R. Dansby
Philip A. Irish 250

The Role of Total Quality Management in Getting the Anti-Personnel
Obstacle Breaching System to Operation Desert Storm

Felipe A. Garcia
Stanley H. Holgate
Carolyn K. Bensei
Steve Jones 254

The Total Quality Management Model

Sue Huang
Jerry Davies
Kay Barrow 255

A Comprehensive System for Minimizing Organizational Delinquency

Robert T. Hogan
Joyce Hogan
Ralph M. Carney
Janice S. Houston
Chase P. McCown
Frederick W. Gibson
Gordon J. Curphy 257

Predicting Organizational Delinquency Among Pre-Screened Subjects: A
Concurrent Validity Study

Gordon J. Curphy
Frederick W. Gibson
Kent A. Landreth
Chase P. McCown 258

	PAGE
Selection in Leadership: Some New Research Directions	
Robert T. Hogan	
Joyce Hogan	
Dianne L. Nilsen	
Richard L. Hughes	
Frederick W. Gibson	
Gordon J. Curphy	263
Equal Opportunity Climate and Its Measurement.	
Dan Landis	
Mickey R. Dansby	
Paul Rosenfield	
Philip A. Irish	
Ronald Joe	264
Medical Reserve Personnel Demobilized After Operation Desert Shield/Storm	
A. David Mangelsdorff	
Shelley Perry	
James Griffith	
Ludwig Uhlmann	
D. Stephen Nice	268
Observations on the Role of Cohesion in Operation Desert Storm	
Frederick Manning	
Robert Gifford	
Kathleen M. Wright	
Paul Bartone	
James Griffith	
Carol Nicholson	
Joe Kulbok	
Beverly C. Harris	
Alma G. Steinberg	281
Operations Desert Shield and Desert Storm: WRAIR's Studies of Soldiers'	
Coping and Adaptation	
David H. Marlowe	
Paul T. Bartone	
Gregory L. Belenky	
Rober K. Gifford	
Kathleen M. Wright	283

	PAGE
Army Family Stress and Coping in Operation Desert Shield/Storm	
Joel M. Teitelbaum	
Sally Bell	
Leora Rosen	
Florence Rosenberg	
Maria Snyder	
John Wells	
David J. Westhuis	287
Author/Presenter Index	288



Multi-Modal Assessment and Treatment of Airsickness

Major R. Jeffrey Jackson, Ph.D.

Abstract

Airsickness is a disruptive phenomenon in military flight training and operational settings. Despite its occurrence, several studies have demonstrated efficacy in treating this syndrome. Although generally successful, these treatment strategies appear to be guided by their presumed effectiveness rather than a theoretical rationale. In support of these varied treatments, a multi-modal conceptual model is proposed. This model attempts to integrate those factors bearing on the etiology of airsickness, and, based on a careful assessment, directs and explains the treatment emphasis.

Airsickness has been and continues to be a deleterious facet of military flying. The overall incidence of airsickness, however, has been difficult to precisely ascertain, and seems to vary according to the type of setting and nature of the survey. Nonetheless, in military flight training airsickness has been reported at rates as low as 10% and as high as 75% (Giles & Lockridge, 1985; Jones, Levy, Gardner, Marsh, & Patterson, 1985; O'Hare & Roscoe, 1990). Thus, it is reasonable to expect that a sizable minority of aspiring aviators will be deterred, to some extent, by symptoms of airsickness.

Although airsickness occurs with a relatively high frequency, it is encouraging that there are a variety of airsickness treatment programs that have reduced or alleviated airsickness symptoms using psychological approaches (e.g., Dobie, May, Fisher, & Bologna, 1989; Giles & Lockridge, 1985; Jones, et al., 1985). Biofeedback, relaxation, and cognitive modification training have all shown efficacy for reducing motion discomfort. In addition to using current psychological strategies, these programs have exposed their subjects to a moving environment. Some used mechanically generated motion sensations in a Barany Chair (Giles & Lockridge, 1985; Jones, et al., 1985) or in an Equinox 2000 (Mallery, Berger, Wittman, Mastroianni, & Schoen, 1991) and one used visually induced motion sensation (Dobie, et al., 1989). Cumulatively, these treatment efforts represent a broad based approach to the problem of airsickness.

These studies provide independent support for using a psychologically based training approach to help student pilots manage their airsickness. While this is certainly important, most of this research concentrates on management techniques and not on developing a better understanding of the etiology of airsickness. In other words, all but the work by Dobie et al. (1989) offer a treatment strategy that is (apparently) directed by an implicit or unarticulated theory. Consequently, any gains in airsickness management must be implemented in a reactive rather than preventive fashion. Treatment without theory does not permit reasonable testing of etiological ideas.

It is perhaps too critical to assert that these clinical outcome studies provide minimal insight into the etiology of airsickness. In fact, one can

92-17646



sift through these studies and draw some reasonable assumptions about the factors that contribute to airsickness. For example, since biofeedback, relaxation, and cognitive training are associated with improvement in airsickness symptoms, it can be inferred that certain sensations, tension behaviors, and disturbing thoughts are contributing to the rise of this syndrome. Furthermore, the improvements associated with exposure to motion stimuli suggest that there is either some physical tolerance (training for motion sensitive neural pathways) or that the motion environment provides a test for recently learned psychological strategies. In any event, it appears that at least four variables are related to the onset and amelioration of airsickness--sensation, behavior, cognition, and physical factors.

By considering a number of factors in the etiology of airsickness, the conceptualization of this syndrome becomes far more advanced than unidimensional frameworks. Thus, it is important to move beyond thinking about airsickness as primarily a result of a neural mismatch. It is even necessary to develop a model that goes beyond the integrated process model articulated by Dobie et al. (1989) in which there is an interaction between cognitive factors and sensory/perceptual activity. Although their model provides a nice account for some of the dynamics of airsickness, it falls short in two respects. First, it does not include some of the immediate factors that are likely to arouse airsickness (e.g., evaluation/performance anxiety in student pilots). Second, since the target of the intervention is just belief surrounding the history of motion sensitivity, it suggests that little can be done to address other factors precipitating airsickness. Nonetheless, it does provide an integrated foundation for a more comprehensive model.

A broader model would significantly address these shortcomings. Such a conceptualization would reflect an integration of the many biopsychosocial factors contributing to the development of airsickness symptoms. This model would allow for a broad band assessment of the problem, and correspondingly, would direct a sweeping, yet systematic strategy for intervention.

Although this may seem to demand a great deal, the framework for such a model has already been developed. Lazarus (1981) proposed a theoretical framework that encompasses these important criteria within his multi-modal therapy approach. Not only does his conceptualization guide a thorough and integrated assessment, it also guides a systematic intervention strategy under the rubric of informed eclecticism. The airsickness model to be proposed here maps directly onto the framework of multi-modal therapy.

In his model, Lazarus (1981) describes seven basic components that are relevant to the onset of airsickness. These components are behavior, affect, sensation, imagery, cognition, interpersonal factors, and biological processes. In Lazarus' acronym, this is the BASIC ID (identity) of problems. A particular pattern of excesses and deficits in each of these areas would be predicted to give rise to airsickness (see Figure 1).

The component or dimension labelled behavior as it relates to airsickness is somewhat narrowly defined. It includes a number of actions a student pilot does prior to or during flight. However, this dimension is not used to describe behaviors having an immediate or direct impact on biological functioning (e.g., eating behavior). Examples of behavioral excesses would include being tense, holding one's breath, dramatic head movements, and

mechanical flying. Behavioral deficits would be marked by inadequate study and limited chair flying.

Affect becomes a factor as a result of the relationship between arousal and performance. As emotional arousal becomes intense, as in anxiety, frustration/anger, excitement, and shame, performance is likely to be compromised. In the aroused affective state there is insufficient satisfaction or other calming emotional experiences.

In contrast to emotional feelings, sensation refers to interoceptive and proprioceptive cues. Most airsick student pilots report excessive feelings of nausea and usually some combination of sweating, a lump in their throat, changes in saliva production, altered breathing pattern, and/or increased heart rate. In general, student pilots experience an intense array of sensory inputs. What they seem to lack is sufficient familiarity with such stimuli. Consequently, they show a deficit in comfort.

Imagery reflects the individual's core assumptions and underlying beliefs. In many cognitive models, this refers to the schema rather than the thought content or process. Imagery can effect airsickness if the student pilot's self view is overly rigid, perfectionistic, and locked into a particular performance level and obtaining a certain type of aircraft in which to professionally fly. The susceptibility for airsickness may be somewhat higher for those who have difficulty accepting themselves as "good enough".

Whereas imagery reflects schema based operations, cognition addresses the automatic content and process of thinking. Cognition may heighten the likelihood of airsickness in individuals whose attention is drawn and concentrated on their physical sensations. It may have the same effect on those who have high levels of anticipatory anxiety as a result of unsupported predictions and for those who are worry prone in general. Very much related to this, the absence of coping self-statements may heighten one's airsickness susceptibility.

Military flight training is an interpersonal process. The presence of another person, particularly if that person is evaluating performance, will have some impact on the student pilot. Evaluation concerns coupled with critical feedback that may be hard to hear are likely to upset the student pilot in some way. Thus, it is necessary to appreciate the relationship between the instructor pilot and the student pilot to adequately assess factors contributing to airsickness.

Finally, drugs or biological processes are very important in the development of airsickness. One of the key contributors to airsickness seems to be vestibular hypersensitivity. Thus, an individual who is sensitive to motion is at increased risk of manifesting airsickness in the flight training environment. Also included as biological factors in the rise of airsickness are dietary patterns, sleep concerns, and general health. Clearly the body is important in this multi-modal assessment of airsickness.

Obviously the mind and body, as broadly described in this model, are critical and joint factors in the development of airsickness symptoms. Not only do each of the multi-modal dimensions influence the threshold of airsickness, but the dimensions are also interactive. That is, these components combine or interact with one another to either increase or decrease the likelihood of developing airsickness. For example, a perfectionistic student pilot who is tense in the cockpit, highly cognizant of those

sensations, and prone to interpret such feelings as indicative of extreme airsickness and subsequent failure is likely to increase the actual susceptibility to this syndrome. Thus, a variety of physical and psychological factors contribute to the etiology of airsickness. It is critical to understand the components that give rise to this condition to develop a better theoretical understanding and treatment program.

The treatment of airsickness symptoms logically follows a careful assessment of the multi-modal components. Each component or factor can be reasonably well managed by a corresponding, rationally selected treatment strategy. There is some overlap in terms of intervention possibilities, but the approaches are suited for that factor and reduce the interactive effect of those components.

In an attempt to reduce symptoms that arise in the behavioral component, the clinician may elect to employ a number of strategies. Relaxation training should decrease tension behaviors, motion experience in the Equinox 2000 could curb excessive head movements and erratic breathing patterns, and chair flying with mastery imagery would act as a form of behavioral rehearsal.

The affective contributions to airsickness can be managed by identifying and accepting feelings, using relaxation and biofeedback training, and implementing graded exposure to increase tolerance and desensitization.

Similar procedures can be used to limit the contributions of different or excessive sensations. Also, reattribution of sensory signals as indicative of motion rather than motion sickness can vitiate these contributions to airsickness.

Imagery based contributions can be challenged by identifying and modifying faulty assumptions. While this is a good theoretical prescription, this is likely to be such an ingrained characteristic that it may not change much in a short-term treatment approach.

The cognitive component may be more responsive to intervention. Addressing and correcting expectations, identifying and countering automatic thoughts, and recognizing and addressing information processing errors are useful techniques under this heading.

A variety of socially based strategies can be used to resolve interpersonal difficulties. For the student pilot who feels that no one else should know their "terrible secret", anti-shame techniques can be used. When the problem is more central to the instructor pilot, communication skills and assertiveness training can promote a more harmonious relationship.

Two basic strategies are employed to manage the drugs/biology component. The first is to minimize vestibular hypersensitivity. Continued flying, the use of a medication patch, and motion stimulation can curtail the effects of such sensitivity. The other general strategy is to promote healthy behavior conducive to freedom from airsickness. This includes, but is not limited to, adequate rest and a good diet.

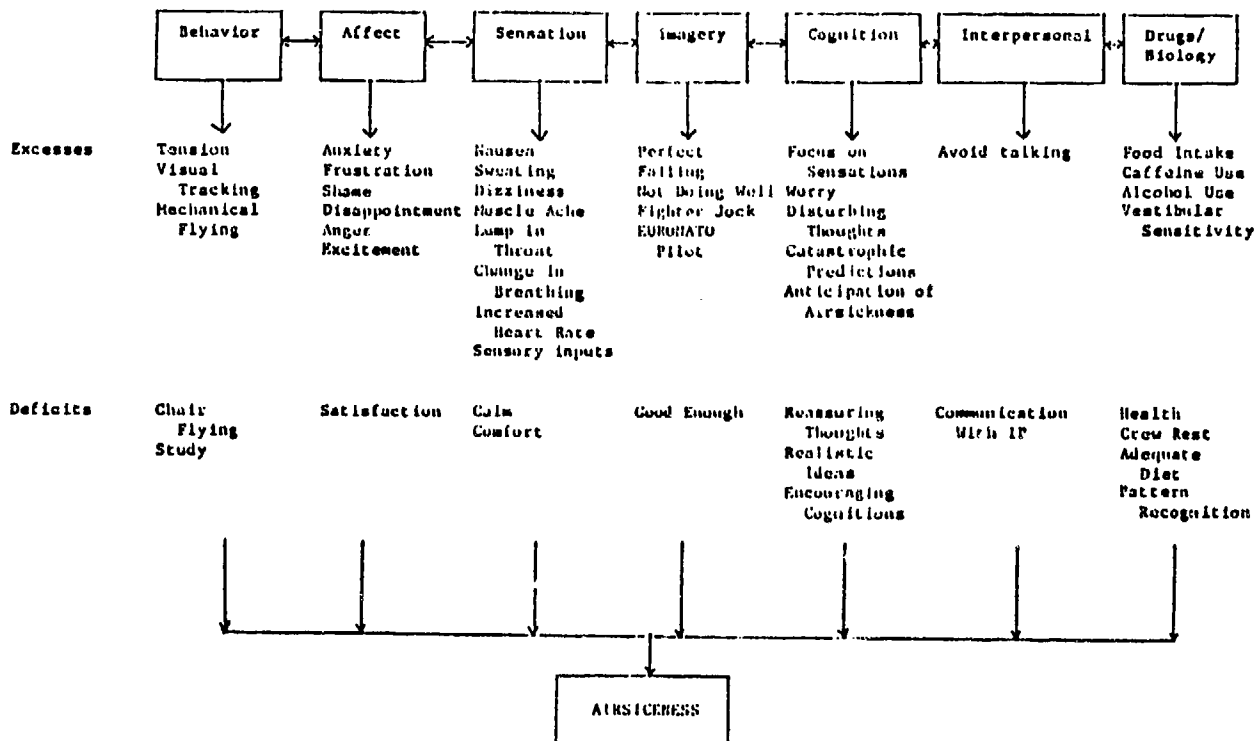
Since previous research has shown effectiveness in treating airsickness using focused clinical strategies, it would be expected that the use of broader or combined strategies would show even greater efficacy. Further, an effective multi-modal treatment would lend validity to the multi-modal conceptualization of airsickness. Such support would add significantly to an understanding of airsickness, and therefore, would allow better control.

References

- Dobie, T.G., May, J.G., Fisher, W. D., & Bologna, N.B. (1989). An evaluation of cognitive-behavioral therapy training for training resistance to visually-induced motion sickness. Aviation, Space, and Environmental Medicine. 60, 307- 314.
- Giles, D.A., & Lochridge, G.K. (1985). Behavioral airsickness management program for student pilots. Aviation, Space, and Environmental Medicine, 56, 991-994.
- Jones, D.R., Levy, R.A., Gardner, L., Marsh, R.W., & Patterson, J.C. (1985). Self-control of psychophysiologic response to motion stress: Using biofeedback to treat airsickness. Aviation, Space, and Environmental Medicine. 56, 1152- 1157.
- Lazarus, A.A. (1981). The practice of multi-modal therapy. New York: McGraw-Hill Book Company.
- Mallery, C.J., Berger, R.C., Wittman, W.T., Mastrolanni, G.R., & Schoen, R.J. (1991). Effects of gyro-fitness training on airsickness management. In Proceedings of the Sixth International Symposium on Aviation Psychology, 752-757.
- O'Hare, D., & Roscoe, S. (1990). Flightdeck performance: The human factor. Ames: Iowa State University Press.

Figure 1

Conceptual Model of Airlsickness Onset





Airsickness Management and the Benefits of Gyro-Fitness Training

Maj William T. Wittman, Maj Carl J. Mallery, Maj Robert C. Berger,
Maj George R. Mastroianni, and Lt Col Robert J. Schoen
United States Air Force Academy

Abstract

This paper examined the role of gyro-fitness training in reducing or eliminating airsickness symptoms in military student pilots. A gyro-fitness training program was developed which allowed airsick cadets to experience motion stimulation and subsequent motion desensitization without being in the actual flying environment. Strong anecdotal evidence suggested the program was helping students overcome airsickness. A formal analysis of the program reported here, suggests training may not be as effective as casual evidence would indicate. This finding has implications for both airsickness treatment and theory.

The Air Force Academy trains approximately 650 cadets a year in the T-41 flight screening program. T-41 is a 15 flight course (including a solo flight) conducted in Cessna 172's. Currently we have a 2% washout rate due to airsickness and more than 50% of the cadets experience some airsickness symptoms during training. Unfortunately, there seems to be no consistently successful screening technique to determine who is prone to airsickness or which individuals will experience the problem only transiently (Dobie, 1989). The Academy has therefore relied on treating individuals as problems occur using an eclectic treatment approach. Student pilots are required to see the Flight Surgeon after the second instance of incapacitating airsickness, that is, after active airsickness or passive airsickness that affects performance. The Flight Surgeon serves as triage, screening for medical reasons, e.g., sinusitis, and adds additional treatments from our counseling staff or spatial awareness training group as necessary. The counseling staff screens for psychological causes which may be aggravating the airsickness problem, e.g., fear-of-failure reactions. Spatial Awareness Training is used to increase a student's experience with self-induced motion. The overall program is similar to the multifaceted approach suggested by Dobie (1989) and (1990), with modifications to the type of medical support used as well as the nature of the training device employed to simulate pitch and roll maneuvers experienced in the aircraft.

In our treatment of cadets, some success has been achieved using motion sickness drugs. Scopolamine, or scopolamine plus dexadrine, and Dilatin have all shown limited success (Erickson, 1990). The T-41 program, however, requires the cadets to be off airsickness medication by the pre-solo ride (11th sortie) thus lessening the time available for drug therapy to work and restricting its usefulness. Experience in flight training has shown that most individuals overcome airsickness as they increase their flying hours. This prompted our use of Spatial Awareness Training to accelerate recovery rates. Using an Equinox 2000 Gyro-Fitness Trainer, a man-sized, two-axis-of-motion gyroscope, each student pilot is able to repeatedly experience abrupt changes in pitch and roll through 360 degrees of motion. The required physical inputs on the Equinox Trainer are unique and induce vestibular accelerations and visual tracking requirements similar to those found in flight. Furthermore.

92-17647



use of such a device is consistent with literature suggesting that some sort of sensory mismatch generally occurs between the vestibular and visual senses, creating airsickness (Igarashi, Kobayashi, Kulecz, & Isogo, 1986; Cynader & Grasse, 1988). Rapid changes in pitch, like those experienced in beginning flight maneuvers such as power-on stalls seem to be the most provocative (Igarashi et al., 1986) and can be readily simulated in the Equinox. But even with this degree of fidelity in the simulation of the induced motion, researchers such as Crowley (1987) have found such training may not transfer to actual flight because factors such as fear associated with the aircraft cannot be reproduced. However, Dobie and May (1990) have found that there are both general and specific components in learning to tolerate motion environments.

This study investigated just this issue; whether Spatial Awareness Training was actually helping student pilots overcome their in-flight airsickness problems. Over the last few years we have provided training to nearly one hundred cadets. Almost all those treated became more tolerant and confident in the Equinox motion environment. More importantly, most completed their T-41 Flight Screening Program. Some believed so strongly in their training they wanted to purchase their own Equinox Trainer. Many suggested that every undergraduate pilot training base should have a trainer on hand. While the anecdotal evidence was encouraging, empirical evidence was needed to determine if of our treatment really was effective and why it was.

Method

Subjects

Subjects were 14 Academy cadets enrolled in the T-41 Flight Screening Program. These cadets were referred to the Spatial Awareness Training Program after two documented cases of airsickness during training flights. Some of these cadets were treated with drugs, some with biofeedback and relaxation training, and some with both forms of therapy before, or concurrently with, enrollment in the Spatial Awareness Program. Since the overriding ambition of most of the cadets enrolled is to fly, those seen in our program were generally highly motivated and often anxious about their performance. For ethical reasons, we could not use the control group of choice, those experiencing airsickness to whom treatment would be denied. We therefore randomly selected 54 cadets (from a total population of 210 cadets in the T-41 program) and tasked them to fill out the same airsickness surveys after each flight that the treated cadets filled out. From this group we selected cadets who were experiencing some airsickness, but not severe enough to require treatment, to serve as a psuedo-control group.

Apparatus

The Equinox 2000 Gyro-Fitness Trainer allows 360 degrees of rotation around two axes. The pitch axis passes through the body from side to side while the roll axis runs from front to back at approximately chest level. By shifting weight and applying pressure to the foot-plate and handgrips appropriately, motion in either of the two axes alone, or complex movements involving both axes simultaneously, can be achieved with some practice. By securing the roll axis the movement of the subject can be limited to the pitch axis. Thus, the device can be configured to permit full freedom of movement in two axes, or can be modified to permit movement in the pitch axis only.

Procedure

Cadets in the program were randomly divided into two groups. The two groups differed in the instructions they were given and the nature of the maneuvers they learned in the Equinox 2000. One group, called the "Flight Analogy" group, used the Equinox in the normal configuration, allowing full freedom of movement in both axes. Cadets assigned to this group progressed through four 30-minute training sessions during which they performed increasingly more complex maneuvers. By the fourth session, cadets in this group were generally quite capable of smoothly executing maneuvers requiring the motion in both axes through 360 degrees of motion. Cadets in this group received frequent feedback on their visual tracking of reference points and were constantly reminded of analogies between the visual and vestibular input they were experiencing in the Equinox and the sensations experienced in actual flight. The remaining cadets were assigned to the "Fitness" group. The main differences between the Fitness group and the Flight Analogy groups were (1) subjects in the Fitness group used the Equinox in the modified configuration, allowing motion in only the "pitch axis (2) subjects in the Fitness group were never allowed to rotate through more than 90 degrees before coming to a stop (3) fitness subjects received no feedback on visual tracking or flight analogy.

Subjects in the two groups were exposed to roughly the same amount of training, but learned a different set of maneuvers and presumably gained very different amounts of both vestibular stimulation and practice in visuo-vestibular coordination. These two groups reflected our assumptions about how we spatial awareness training works -- cadets desensitize to motion, learn how to properly coordinate eye tracking and head movements, and gain confidence that what they do in the Equinox will help them in the T-41. Given these benefits, we assumed the Flight Analogy group would show a treatment advantage. Each cadet in the program was assigned a particular instructor who worked with the cadet for all four sessions. Treatment took place on alternate days (excluding weekends) at the same time of day until the program was completed.

Results and Discussion

Of the 14 subjects in our treatment program, 9 were used in our analysis, 5 from the Fitness group and 4 from the Analogy group. The remaining 5 subjects did not fully complete our symptom survey and could not be included in our analysis. From the 54 students selected to complete the symptom survey, we received 21 complete sets of surveys. In order to be considered complete there could not be more than one missing survey out of the first eight flights. We selected a pseudo-control group of 9 subjects from these 21 students for comparison with the two treatment groups. Students were included in the group if they reported experiencing either nausea, vomiting, or dry heaves during any of their flights. Figure 1 reports the mean number of symptoms for our two treatment groups and the pseudo-control group. Both treatment groups appear to have a greater number of symptoms than the control group for the first two training flights. By the third flight, when most of the treatment subjects were just beginning training, their symptoms decreased to roughly the level of the control group. This reduction holds for the remainder of the training program. Also important in these data was the lack

of a clear difference between the Fitness and Analogy groups. Contrary to our expectations, both groups showed decreases in symptoms despite out treatment manipulations. Because our sample sizes were extremely small, due in part to the small number of cadets who actually receive treatment and to the lack of complete symptom profiles, we did not use inferential statistics in our analysis.

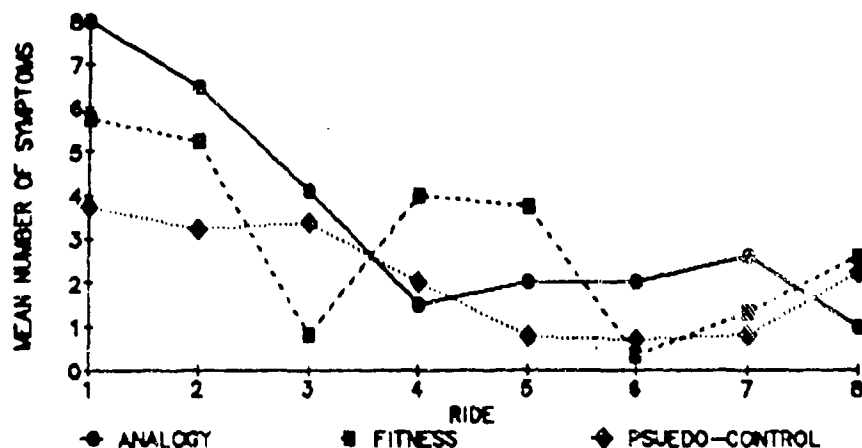


Figure 1. Symptoms by Ride for All Groups

Table 2 provides a profile of the 14 airsickness symptoms we had cadets rate in their symptom surveys. These data reflect the ratings of 26 students who accurately completed at least 7 out of 8 surveys. Subjects from the two treatment groups, the psuedo-control group, and from the original 54 randomly selected subjects were included in the profile. The most commonly experienced symptoms were sweating, burping, and nausea. Also evident is a general decrease in airsickness symptoms across flights, though there are some irregularities in this trend. Weather-related delays between flights, typical in the Spring flying schedule, are partly to blame for these.

	101	201	202	203	301	302	303	401	AVG
BURPING	24%	27%	23%	23%	24%	8%	21%	9%	19.9%
NAUSEA	32%	23%	19%	23%	4%	8%	4%	13%	15.8%
NUMBNESS/TINGLING	4%	4%	4%	0%	4%	0%	0%	0%	2.0%
SWEATY	52%	31%	27%	38%	24%	8%	8%	17%	25.6%
VOMITING	16%	8%	8%	12%	4%	4%	0%	4%	7.0%
LIGHTHEADED	16%	12%	12%	15%	4%	0%	4%	4%	8.4%
YAWNING	12%	15%	19%	12%	4%	4%	8%	9%	10.4%
CLAMMINESS	12%	4%	8%	8%	4%	0%	0%	4%	5.0%
STOMACH TIGHTNESS	20%	23%	12%	15%	8%	8%	8%	9%	12.9%
DISTRACTED	24%	4%	19%	15%	4%	0%	13%	4%	10.4%
DISORIENTED	8%	8%	8%	4%	4%	0%	0%	4%	4.3%
LEG CRAMPS	0%	0%	0%	0%	4%	0%	0%	0%	0.5%
DRY HEAVES	12%	0%	4%	0%	4%	0%	0%	0%	2.5%
HEADACHE	16%	4%	12%	8%	4%	4%	4%	0%	6.5%

Table 1. Percentage of Cadets with Various Aisickness Symptoms.

Unfortunately our results do not allow us to conclude our treatment conditions were different, nor can we conclude our gyro-fitness program helped cadets overcome their airsickness problems. Possibly the most influential factor in recovery from airsickness was exposure to the flying environment. In time, almost all cadets have reduced symptoms. On the other hand, we can't overlook the fact that all the subjects in our treatment groups had seen a Flight surgeon and counselor prior to beginning Spatial Awareness Training. Some were already on medication when our treatment began and others had received biofeedback training. All subjects had been individually counseled concerning their airsickness by some military progressional. One could therefore argue that our results support an eclectic or multifaceted treatment approach such as that suggested by Dobie (1989, 1990). By providing medical treatment, counseling and Spatial Awareness Training, we do help student pilots. Even those with more severe symptom profiles are able to rapidly manage their airsickness difficulties and successfully complete preliminary flying training.

References

- Crowley, J. S. (1987), Simulation sickness; A problem for Army aviation. Aviation, Space and Environmental Medicine, 58, 355-357
- Cynader, M. S. & Grasse, k. (1988). Neural mechanisms underlying visual-vestibular interactions. (DCIEM contract #01 SE 977 11 4 7996). Halifax, Nova Scotia: Dalhousie University, Dept of Psychology.
- Dobie, T. G. (1989). 1988 Smith W. Ames Lecture: Teaching the right stuff -- the heart of the matter. Aviation, Space and Environmental Medicine, 60, 195-196.
- Dobie, T. G. (1990). Motion sickness prevention: A course of instruction in cognitive-behavioral counseling. Proceedings of the 12th Symposium Psychology in the Department of Defense (pp. 87-91). Springfield, VA: National Technical Information Service, U.S. Dept. of Commerce.
- Dobie, T.G. & May, J.G. (1990). Generalization of Tolerance to Motion Environments. Aviation, Space and Environmental Medicine, 61, 707-711.
- Erickson, D. (1990). Sick of work. Scientific American, 262, 84-86.
- Lgarashi, M., Kobayashi, K., Kulecz, W.B., & Isago, H. (1986). Vestibular-visual conflict in pitch and yaw planes in the squirrel monkey. Aviation, Space and Environmental Medicine, 61, 1071-1074.



Flight Simulator-Induced Sickness and Visual Systems Evaluation

Harold D. Warner*
Gary L. Serfoss, Lt USAF**
Todd M. Baruch**
David C. Hubbard*

*University of Dayton Research Institute, Williams AFB, AZ
**Armstrong Laboratory, Aircrew Training Research Division,
Williams AFB, AZ

Abstract

This investigation was conducted to compare the incidence and severity of simulator sickness between two flight simulator visual systems and two groups of pilots. Both visual systems were found to produce an increase in simulator sickness symptomology, and some pilots terminated the simulator sessions due to severe discomfort. Of the pilots who completed the simulator sessions, simulator sickness symptomatology significantly increased over the sessions. There were no differences in the measures of simulator sickness between the two visual systems or groups. The pilots recovered from the adverse effects of the simulation within 30 minutes following the sessions.

Simulator-induced sickness has been found to occur in conjunction with the use of various military flight trainers (e.g., Kennedy, Lilienthal, Berbaum, Baltzley, & McCauley, 1989). The percentage of users that have experienced simulator sickness symptomatology have ranged from 11 to 88 percent, depending upon the simulator (Casali & Frank, 1988). The symptoms of simulator sickness that were observed included: eyestrain, headaches, dizziness, sweating, drowsiness, and nausea.

The Aircrew Training Research Division of the Armstrong Laboratory (AL/HRA) located at Williams Air Force Base, Arizona, is the host facility for two wide-screen flight simulator visual display systems that provide full-color, computer-generated imagery. They are the Limited Field-of-View Dome (LFOVD) and the Display for Advanced Research and Training (DART).

After the DART was installed in 1988, some pilots experienced simulator sickness symptomatology during an initial checkout of the simulator. Similar maneuvers as those performed in the DART have been accomplished in the LFOVD without evoking simulator sickness. Therefore, it was hypothesized that the different simulator technologies employed contributed to the difference in incidence of simulator sickness. The present investigation was designed to test this hypothesis.

Method

Subjects

Twenty-four pilots participated, 16 active duty U.S. Air Force T-37 and T-38 instructor pilots (IPs) and eight "older" pilots and former pilots, who were not currently flying military aircraft. The average age of the IPs was 28.6 yrs, and 52.1 yrs for the older pilots.

Visual Simulation Systems

The DART provides a mosaic of eight pentagonal rear-projection windows surrounding the simulator cockpit. Computer-generated imagery is projected onto the windows. The LFOVD employs a 24-foot-diameter dome, and is equipped with a moveable, head-slaved area-of-interest (AOI) display, which is a high-resolution inset that is surrounded by a larger background display with lower resolution.

92-17648



The visual imagery for both simulators was produced with a General Electric Advanced Visual Technology System (AVTS) computer image generator, which provided a full-color rendition of the visual scene. The cockpit enclosed within each simulator represented an F-16 fighter aircraft, with all the appropriate flight controls and instruments. Neither simulator was equipped with a motion base.

Visual Environment

The simulated visual environment consisted of a continuous canyon with a narrow floor, high, sloping walls, and a cloud layer above. Numerous lakes, farms, fields, and other two-dimensional textured shapes were present, as well as a wide variety of three-dimensional objects such as trees, vehicles, and buildings. The same visual environment was used with both the LFOVD and DART visual simulation systems.

Performance Task

The task consisted of: (a) a 5-minute simulator familiarization flight, (b) a 5-minute practice formation flight, (c) a 20-minute, single-ship flight, and (d) a 20-minute formation flight. In the familiarization and single-ship flights, the pilots were instructed to bracket their airspeed between 425 and 475 knots and stay below 500 feet AGL (above ground level). In addition, the pilots were requested to follow the road in the canyon for the single-ship flight and verbally signal the presence of the vehicles on the road.

One prerecorded lead aircraft flight was used for all formation flights. Pilots were instructed to fly a combat trail formation with a slant range separation of 2,000 feet or less.

Simulator Sickness Assessment

Simulator sickness was assessed using a simulator sickness questionnaire (SSQ) and two ataxia (i.e., postural equilibrium) tests, which have been employed in other simulator sickness evaluations (Kennedy, et al., 1988; McCauley, Hettinger, Sharkey, & Sinacori, 1990). The SSQ is comprised of a symptom checklist that permits a determination of the symptoms the pilots experienced and the severity of the symptoms. The ataxia tests were: Stand on Leg Eyes Closed (SOLEC) and Walk on Floor Eyes Closed (WOFEC). The SSQ and ataxia tests were administered immediately prior to and following the simulator sessions, and again 30 minutes after the sessions. In the SOLEC test, the time subjects were able to stand on one leg, for up to 30 seconds, was recorded three times for each of the pre, post 1 and post 2 administrations. For the WOFEC test, the number of steps, up to 12, the subjects were able to make heel-to-toe on a line were recorded, and the test was repeated three times for each time of administration.

Pilot discomfort ratings were also obtained from the pilots at 5-minute intervals during the simulator flights. A 7-point scale was adopted, where 1 indicated "normal, symptom free" and 7 indicated "severe discomfort."

Procedure

Each pilot participated in two simulator sessions, one with the LFOVD and one with the DART. The order of sessions was counterbalanced, and the sessions were at least two weeks apart. Within each session, the pilots first performed the 5-minute simulator familiarization flight, then the 5-minute practice formation flight, and then a counterbalanced combination of the 20-minute, single-ship flight and the 20-minute, formation flight.

Results

Session Terminations

Due to severe discomfort, one IP and two pilots in the older group were unable to complete the simulator sessions with either the DART or LFOVD. Also, one of the pilots in the older group terminated the session when the DART was used, but was able to complete the session with the LFOVD. At the conclusion of the latter session, however, this subject exhibited severe sweating and pallor, which are prominent overt symptoms of simulator sickness. Overall, 16.67% of the pilots terminated the sessions with the DART because of severe discomfort, and 12.5% were unable to complete the flights with the LFOVD.

Pilot Discomfort Ratings

The pilots who terminated the sessions due to severe discomfort discontinued the simulation at different time periods. The mean discomfort rating after the first five minutes for these pilots was 2.36, and the mean rating at the end of the five-minute period just prior to the pilot terminating the sessions was 4.14. Since this latter mean is well below the maximum scale value of seven, it is evident that there may have been a very precipitous increase in discomfort from the final rating to the termination of the session. It is also possible that the pilots were reluctant to report high discomfort ratings. To determine whether the pilots who were able to complete the sessions experienced an increase in discomfort over time, the pilot discomfort ratings for these pilots were subjected to an analysis of variance (ANOVA) with repeated measures. The analysis showed that the main effect of time was significant, $F(9,162) = 2.97$, $p < 0.01$. The mean discomfort ratings increased from 1.24 for the first rating to 1.73 for the final rating. The mean discomfort ratings are plotted in Figure 1. The ratings increased on the average up to the 30-minute mark where the task was changed. The ratings declined at this point, but then increased to an even higher mean level. The main effects of visual system (LFOVD vs. DART) and pilot group (IPs vs. "older" pilots) were not significant, nor were any interactions.

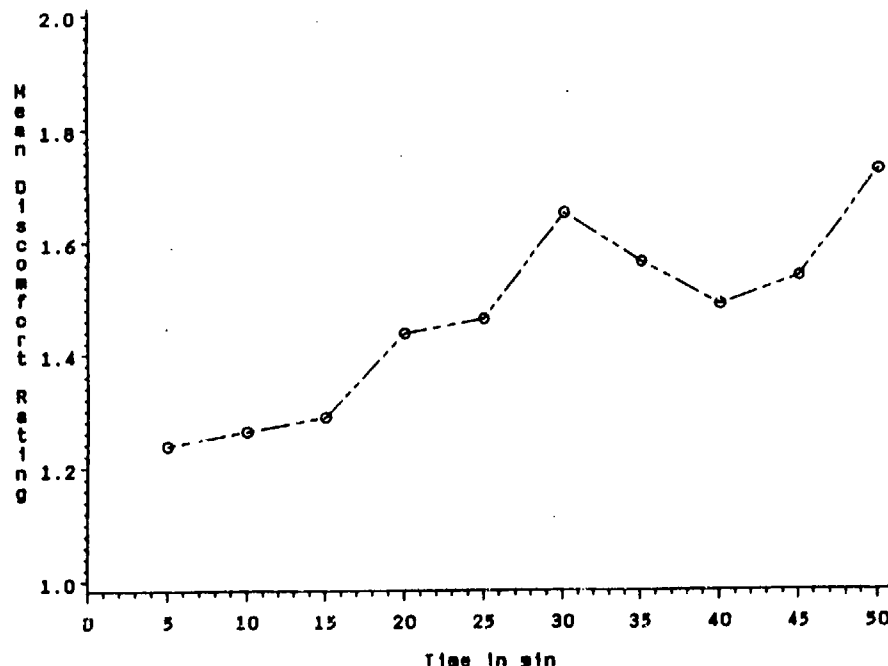


Figure 1. Mean pilot discomfort ratings, excluding those too ill to finish.

Simulator Sickness Questionnaire (SSQ) Analysis

The SSQ data collected in relation to the pilots who prematurely exited the simulators were not considered an accurate reflection of the pilots' symptoms because they were allowed to recover from the adverse effects of the simulation before the post-session SSQs were administered. For this reason, only the data from the pilots that completed the sessions were analyzed. The SSQ data were transformed into a single, overall score using the method developed by Lane and Kennedy (1988). Scores were produced for the pre, post 1, and post 2 SSQ checklists for each of the pilots. An ANOVA with repeated measures was used to analyze the transformed SSQ scores. A significant main effect of time of administration was obtained, $F(1,72) = 10.20$, $p < 0.01$. The mean transformed SSQ scores for the three successive time periods were: 105.88, 117.67, and 103.93. The main effects of visual system (LFOVD vs. DART) and pilot group (IPs vs. "older" pilots) were not significant, nor were any interactions. Of the pilots who were able to complete the sessions, 75% reported an increase in at least one symptom on the post 1 SSQ following the flights with the DART, and 60% reported an increase after the session using the LFOVD.

WOFEC and SOLEC Analyses

Post-session WOFEC and SOLEC tests were not administered to the pilots who were unable to complete the simulator sessions because the tests could not be administered immediately following the sessions as a result of the pilots' discomfort. Therefore, the post-session WOFEC and SOLEC tests were administered only to the pilots that completed the sessions. An ANOVA with repeated measures was used to analyze the ataxia data. A significant main effect of time of administration was observed for the WOFEC test, $F(2,94) = 7.15$, $p < 0.01$. The mean WOFEC scores for the pre, post 1, and post 2 tests were 6.19, 4.78, and 5.89. The group and visual system main effects were not significant, nor any interactions. The analysis of the SOLEC data indicated that not any of the main effects or interactions were significant.

Discussion

It was shown that both the LFOVD and DART visual systems can induce simulator sickness symptomatology. Based on the number of pilots who terminated the sessions, as well as the various measures of simulator sickness that were used, the incidence and severity of the symptoms produced were approximately the same for both systems.

For the most part, the same pilots who terminated the session using one system terminated the session with the other system. One pilot, who prematurely terminated the DART session, was able to complete the LFOVD session, but experienced severe symptoms of simulator sickness in conjunction with the use of the LFOVD. Although the sample size was limited, a higher percentage of the "older" pilots terminated the sessions than the IPs.

The analyses further revealed that there was an increase in simulator sickness symptomatology within the sessions among the pilots who were able to complete the sessions. First, the self reports of pilot discomfort increased slightly, but significantly, over time. Next, the incidence and severity of symptoms, as indicated by the SSQ, significantly increased between the pre and post 1 administrations. Finally, postural disequilibrium, as measured by the WOFEC test, increased significantly between the pre and post 1 tests. The changes in symptomatology over time were essentially the same with both visual systems and for both groups of subjects.

Thirty minutes following the simulator sessions the pilots who completed the flights recovered from the adverse effects of the simulation. The mean SSQ scores associated with the post 2 administration was nearly the same as the pre-flight mean. In addition, the post 2 mean WOFEC score was almost equal to the pre-flight mean.

References

- Casali, J. G., & Frank, L. H. (1988). Manifestation of visual/vestibular disruption in simulators: Severity and empirical measurement of symptomatology. Proceedings of the Conference on Motion Cues in Flight Simulation and Simulator Induced Sickness (AGARD-CP-433). Neuilly Sur Seine, France: North Atlantic Treaty Organization, Advisory Group for Aerospace Research and Development.
- Kennedy, R. S., Lilienthal, M. G., Berbaum, K. S., Baltzley, D. R., & McCauley, M. E. (1989). Simulator sickness in U.S. Navy flight simulators. Aviation, Space, and Environmental Medicine, 60, 10-16.
- Lane, N. E., & Kennedy, R. S. (1988). A New Method for Quantifying Simulator Sickness: Development and Application of the Simulator Sickness Questionnaire (SSQ) (EOTR 88-7). Orlando, FL: Essex Corporation.
- McCauley, M. E., Hettinger, L. J., Sharkey, T. J., & Sinacori, J. B. (1990). The effects of simulator visual-motion asynchrony of simulator induced sickness. AIAA Flight Simulation Technologies Conference and Exhibit, Sept. 17-19, 1990, Dayton, OH (AIAA-90-3172). Washington D.C.: American Institute of Aeronautics and Astronautics.



APPLYING TOTAL QUALITY TO A STATISTICS
AND RESEARCH DESIGN COURSE:
A CASE IN POINTS

92-17649



John Micalizzi, Ph.D
David Porter, DPhil
USAF Academy

Abstract

The application of Total Quality (TQ) to classroom learning was illustrated using a statistics and research design course taught at the USAF Academy. Two guidelines, in particular, were used to help develop the structure of the course in the Fall 1991 semester: "drive out fear" and "empower the customer". Specific techniques for implementing these guidelines were presented and discussed. Student satisfaction and performance data were collected to assess the effectiveness of this approach. While the results were mixed, the potential for improving classroom learning clearly warrants continued investigation.

All students majoring in Behavioral Sciences and Leadership at the USAF Academy are required to enroll in a two semester sequence of courses in statistics and experimental design. The initial course, Behavioral Science 331, is taught in the fall semester and focuses on the fundamental principles of descriptive and inferential statistics from the one-sample t-test to One-Way Analysis of Variance. Behavioral Sciences 332, taught during the following spring semester, deals with general issues of scientific research as well as advanced statistical procedures such as multiple regression and analysis of covariance. Together, these courses provide a solid academic foundation in research methodology and problem solving.

In the past, instructors teaching these courses have expressed concerns about both the quantity and quality of student learning. In particular, two issues have consistently surfaced: (1) Some students attracted to the Behavioral Sciences major may be attempting to escape from the mathematical and engineering orientation of the core curriculum. The past academic experiences of some cadets consist of marginal performance in several quantitative courses. As a result, both their actual abilities as well as their perception of these abilities is low. In the most extreme cases, this might be referred to as "statistophobia". (2) Based on their previous experience in Introductory Psychology, most students do not expect extensive quantitative requirements in the major. Coercive approaches, similar to those some students have experienced in traditional mathematics and engineering courses only seem to exacerbate the problem. Impersonal instruction and the use of fear to force students to learn the material are almost always counterproductive in the long run (Glasser, 1990).

The Total Quality (TQ) Alternative

These conditions created fertile ground for the application of TQ principles (Porter, 1991). Two principles, in particular, were used to guide the preparation of Behavioral Science 331 in the fall of 1991. The first,

"drive out fear" (Aguayo, 1990), was aimed at reducing the threat of statistics for those students who may have experienced little academic success in previous quantitative courses. The second guiding principle, "empowering the customer," was used to give students both the opportunity and responsibility for continually improving their academic performance (Lam, Watson, & Schmidt, 1991). Together, these principles helped create a learning environment where both teachers and students jointly pursued knowledge and understanding of behavioral science research methods.

Reducing the threat of statistics involved confronting student's adversarial feelings at three levels: (1) among the students themselves; (2) between students and the teacher; and finally, (3) between students and the course material. Obviously, these feelings interact; student feelings about the course carry over to his/her feelings about the teacher. It should be noted that feelings of fear on the part of students are likely to preclude any actual learning from taking place (Porter, 1991). In the language of Total Quality, time and effort which do not contribute to the intended product (viz., student learning) are simply "waste." However, replacing feelings of competition and suspicion with feelings of cooperation and trust, students learn more and also increase their enjoyment of the class in particular and attitude toward academics in general.

When competition is fostered in the classroom, students focus on the accumulation of points and relative rankings and are less concerned with really understanding the material or forming positive relations with classmates. The intrinsic motivation to learn is supplanted by the extrinsic incentives to get a high grade (Deci, 1980). We tried a two prong attack to solve this problem by eliminating even the inference of "curve" grading and by also basing a significant portion of their course grade on team performance. The criterion-based grading system helped students realize that their performance was a direct result of their own efforts and not dependent on other students doing poorly. Classmates were now looked upon as potential sources of knowledge and emotional support rather than adversaries competing for a limited number of good grades. In addition, small, cooperative teams of students were formed in each section to decide on, prepare and present research projects worth 40% of their final grades. Two-thirds of their grade on these projects was based on the combined effort of the entire team. As a result, students had an incentive to help team members prepare the best possible project. Minimizing conflicts between students resulted in a more positive and nurturing classroom environment.

A similar approach was also used to reduce tensions associated with teacher - student relationships. Up to a point, narrowing the gap between the student and teacher can enhance the cooperative learning environment by changing the role of the teacher from purely an evaluator to more of a mentor. The interactions between students and mentor-based teachers are both more frequent and more positive. Individual personalities and abilities of both the teacher and student affect the development of their relationship, however, the course structure also can either encourage or discourage the tendency to form "learning coalitions". In Behavioral Science 331, responsibility for teaching and conducting classroom activities was decentralized whenever possible. Course-wide learning objectives kept all instructors on the same general track, but plenty of room was left for individual instructors to modify and improve learning in their classrooms.

Reducing the threat of the course material itself was probably the most important issue to deal with because these feelings also spilled over on students' feelings toward the teacher, fellow students and the learning environment in general. While maintaining high but achievable learning goals, we tried to give students many different opportunities to succeed. Most students have preconceived notions of their own relative academic strengths and weaknesses. By providing a variety of different evaluation methods, students felt less threatened that their particular weaknesses would be exploited by any single method of testing. A student's final grade was based on a variety of graded material including in-class tests consisting of multiple choice, short answer and essay/problem solving, group projects, individual contributions to group efforts, optional computer exercises, extra credit papers, and individual and group oral presentations.

Once the fear of failure was minimized, we attempted to turn the responsibility for learning over to students whenever possible. However, we also quickly realized that it was important to accompany increased responsibility with the means to succeed. Students should feel in control of their academic performance and be aware of the various opportunities which exist to continually improve their grade (Dweck & Leggett, 1988).

All students received detailed syllabi and lesson learning objectives. The syllabus specified course goals, student evaluation methods and criteria, and a lesson by lesson breakdown of assignments. In addition, operational definitions of letter grades were included to inform students that their research projects would be evaluated on the basis of specific quality requirements. The purpose of the syllabus was to increase student awareness of the content and process of the course at the very beginning of the semester. Lesson learning objectives also served to reduce student uncertainty by clearly specifying the knowledge and skills which would be tested on the evaluations. Taken together, these course materials provided the information and the necessary means for students to take charge, and therefore take responsibility, for their academic performance.

Once immersed in the course process, students quickly realized that the emphasis in this course was not on memorizing formulas or procedures, but rather developing the ability to apply a general problem solving approach using statistics as a tool for analysis. Taking good notes and depending on the teacher, exclusively, for the right answers was not sufficient to assure high grades. Students needed to participate in class demonstrations and depend on each other to become successful. Again, responsibility for learning was shifted away from the instructor and back to the students in the class.

Students were also allowed several opportunities to choose their tasks. Students, as a team, selected the topic for their research projects. On an individual basis, they could also choose to take advantage of optional computer exercises or extra credit papers to earn additional points in the course. Students were also reinforced by additional bonus points on tests for superior group performance. For consistently outstanding individual performance on tests, validation of the final exam was possible. By giving students a wide range of choices for succeeding in the course, they truly believed that their academic success (or failure) was in their own hands.

Results

The successful application of Total Quality to course structure didn't stop here. Periodic assessments of our progress were needed to make adjustments and continually improve the quality of learning. Data from both student critiques and course evaluations were collected to ascertain the effects of this approach on student learning and satisfaction. Student critique data was used to measure student satisfaction based on several important elements of the learning environment: fellow classmates, textbook, research projects, exams, the overall course, and the instructor. The course-wide results were mixed. Changes in satisfaction between the two semesters did not show dramatic improvements in student perceptions. However, since many of these changes were evolutionary refinements of an already successful approach, it might be unrealistic to expect large-scale improvements in only one semester. There is also the difficulty caused by the fact that two of the three experienced instructors were replaced by two new instructors who had not taught the course for several years. The mean ratings for the only instructor who taught the course both semesters (Fall 90 and 91) clearly improved across all course elements. The fact that the improvements in "the instructor" category for the data from a single instructor (Figure 1) shows the least improvement, argues against the instructor rather than the course design being the primary cause of improvement. Data from student performance across the two semesters also reflected continuous improvement. Using nearly identical criteria, the course GPA increased from 3.18 in the Fall 1990 semester to 3.41 for the Fall 1991 semester. This increase was also accompanied by a 22% decrease the variance between students ('90, $sd^2 = 0.40$; '91, $sd^2 = .31$). The restricted range of scores in the upper end of the distribution may have contributed to the reduction in variance; however, the overall pattern seems to be a general increase in academic performance across all ability levels using virtually the same evaluation instruments.

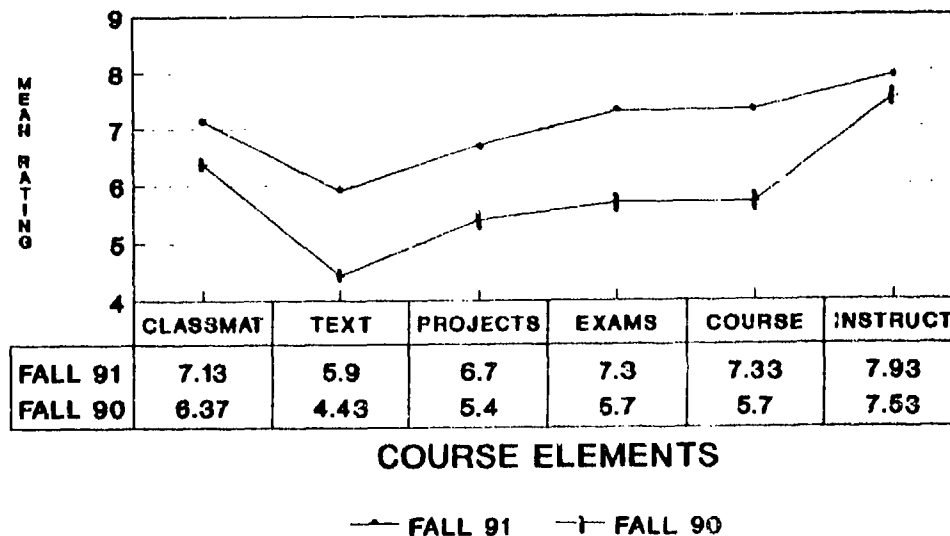


Figure 1. Mean student critique ratings for Beh Sci 331. (instructor)

Naturally, new ideas, better assessment methods and a complete commitment to continuous improvement will be needed to exploit the advantages of applying Total Quality principles to education. We've taken our first tentative step toward understanding how Total Quality can enhance classroom learning, but there remain many more along the way.

Note: Views expressed in this article are those of the authors and do not necessarily reflect the views of the USAF Academy nor any other government agency.

References

- Aguayo, R. (1990). Dr. Deming: The American who taught the Japanese about quality. New York: Carol Publishing Group.
- Deci, E.L. (1980). The psychology of self-determination. Lexington, MA: Lexington Books.
- Dweck, C.S. & Leggett, E.L. (1988). A social-cognitive approach to motivation and personality. Psychological Review, 95. pp. 256-273.
- Glasser, W. (1990). The Quality School: Managing students without coercion. New York: Harper and Row.
- Lam, K.D., Watson, F.D., & Schmidt, S.R. (1991). Total Quality: A textbook of strategic quality leadership and planning. Colorado Springs, CO: Air Academy Press.
- Porter, D.D. (1991). A perspective on college learning. Journal of College Reading and Learning, 24(1). pp. 1-15.



A SOFT-SYSTEMS MODEL FOR CONTINUOUS IMPROVEMENT

Andrew G. Stricker, Ph.D.
Thomas McCloy, Ph.D.
US Air Force Academy

92-17650



Abstract

This paper outlines a model to continuously improve a soft system. The authors take a systemic approach in the examination of soft systems. A systemic view emphasizes connections within and between systems and processes of communication and control (Archibald, 1989). We use the term "soft system" to refer to systems which involve people and machines (Checkland, 1981). The approach we outline in this paper illustrates a multi-causal model for improving soft systems. It is unlike a uni-causal model which would treat practical problems like technical problems. The model we suggest in this paper offers tools to examine practical problems from multiple perspectives. Furthermore, the model offers considerable advantages over more traditional experimental-laboratory models which are exceedingly difficult, if not impossible, to apply outside a controlled laboratory environment.

Under classical experimental models, involving full factorial designs, it becomes nearly impossible to examine complex soft systems. Attempts to do so have involved taking apart the system and analyzing components in manageable chunks and then reassembling the separate analyses to construct a complete picture of the system. This is problematic because such analyses are done under artificial circumstances and it is nearly impossible to examine how components of the system interact with all criteria of interest. The model depicted in Figure 1 combines Taguchi designs with soft-systems thinking to offer the user the means to understand how the system operates and with it the tools to continuously improve operations.

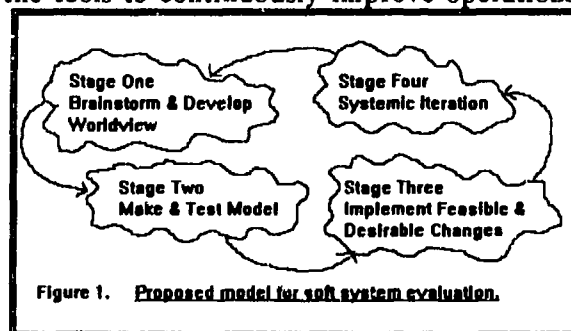


Figure 1. Proposed model for soft system evaluation.

The model shown above takes advantage of fractional factorials and total quality tools to provide understanding of the system necessary for making improvements. Stages in the model are discussed below.

Scenario

A computer company, Nexus Inc., wants to improve its chip manufacturing process. A work team of production supervisors, and selected key workers, has been asked to come up with a way to improve the defective chip identification process.

Stage One: Brainstorm & Develop World View

The work team should find out about the problem situation (ways to improve the defective chip identification process) without trying to impose a structure on it. The intent is to build up the richest possible picture of the situation being studied. This would involve collecting as many perspectives of the problem as possible from a wide

range of people in the problem situation (see Senge, 1990). Particular attention should be focused on elements of structure and process. The relationship between these components create the situational climate or "bound the problem." Elements of structure entail such items as physical layout, reporting structure, and pattern of communication. Likewise, process elements reflect how processes are monitored and provide means for corrective action.

Next, the team develops a common "world view" of the system such that constraints or boundaries are placed on the "structures" and "processes," making it possible to test them via a form of sensitivity analysis. In our illustration, the team's common "world view" of the system is one in which the company minimizes the time it takes to detect defective computer chips.

Stage Two: Make & Test Model

Having agreed on a "world view," the team now must describe what the system has to do to fulfill the vision articulated in stage 1. To proceed, the team views the system from the perspective of : Input ----> Transform ----> Output. Various inputs, such as, raw materials and skill levels of workers, go into efforts to produce quality chips. Transformations of the input into output involve processes associated with production.

The work team decided that one way to improve quality would be to minimize the amount of time it takes for human assemblers to identify defective chips (this would provide more time for other assembling activities). They examined the processes associated with inspecting chips and determined that they could control three factors: (1) scanning techniques used for defective chip identification; (2) whether or not feedback was provided to the human inspectors; and (3) size of chip batches undergoing inspection.

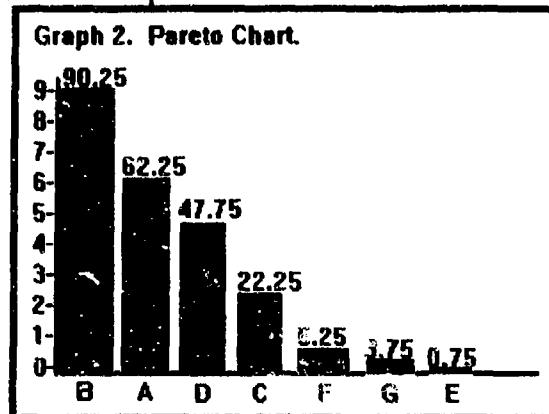
The Model Is Tested . The team now has to determine which settings (or levels) are best for each of the three factors. Factor A (scanning technique) has two possible settings: color or numeric scan code. Factor B (feedback) has two settings: off or on. And factor C (batch size) has two settings: 6 or 8. A Taguchi-L8 design was selected to minimize production runs of the inspection process and to insure orthogonality of all possible combinations of three factors across all settings (readers are referred to Schmidt & Launsby, 1989; and Kiemele & Launsby, 1991). The team decided that inspector response time was the criterion of interest. Furthermore, the response time could be decomposed into two elements: decision and motor time. Ultimately, the team decided to measure decision time. The lower specification limit (LSL) was set for 400 msec and the upper specification limit (USL) was set for 800 msec.

Table 1. Illustrated Block Design with Response Values (msec).				
Feedback				
B ₁ (no=low=1)		B ₂ (yes=high=2)		
Chip Batch Size				
C ₁ (6=low=1)		C ₂ (8=high=2)	C ₁ (6=low=1)	C ₂ (8=high=2)
A ₁ (off=low=1)	Run 1	Run 2	Run 3	Run 4
	y1 794	y1 850	y1 530	y1 550
	y2 810	y2 880	y2 540	y2 565
	y3 820	y3 840	y3 535	y3 620
Scan Filter	Run 5	Run 6	Run 7	Run 8
	y1 596	y1 640	y1 620	y1 540
	y2 580	y2 635	y2 515	y2 550
	y3 570	y3 655	y3 510	y3 530
A ₂ (on=high=2)	Run 5	Run 6	Run 7	Run 8
	y1 596	y1 640	y1 620	y1 540
	y2 580	y2 635	y2 515	y2 550
	y3 570	y3 655	y3 510	y3 530

Table 1 shows the block design with the collected decision times (response values). Table 2 shows the Taguchi-L8 design with the response values from 24 assemblers. In the spirit of keeping our simulation simple we've only used 24 responses.

Table 2. Taguchi-L8 Matrix with Computations.														
Run	A	B	A*B	C	A*C	B*C	A*B*C	Response value $\bar{y}$			s^2	S/N		
1	1	1	1	1	1	1	1	794	810	820	808	171.8	-58.14	
2	1	1	1	2	2	2	2	850	880	840	857	433.0	-58.65	
3	1	2	2	1	1	2	2	530	540	535	535	26.0	-54.56	
4	1	2	2	2	2	1	1	550	565	620	578	136.7	-55.25	
5	2	1	2	1	2	1	2	596	580	570	582	171.8	-55.29	
6	2	1	2	2	1	2	1	640	635	655	643	108.3	-56.16	
7	2	2	1	1	2	2	1	520	515	510	515	26.0	-54.23*	
8	2	2	1	2	1	1	2	540	550	530	540	100.0	-54.64	
AVG(H)								694	722	680	610	631.5	627	636
AVG(L)								570	542	584	654	633	637	628
SS _{effect}								124	-180	-95.5	44	1.5	10.5	7.5
F _{effect}								310.9	853.6	82.8	39.7	0.46	2.21	1.12
									</					

Finally, we constructed a pareto chart (Graph 2) to depict the order of factor magnitude. We converted the absolute difference between factor marginal means using a factor of 2 just to keep the chart simple.



Examination of Graphs 1 and 2 reveal that all three primary factors (B, A, and C) contribute to noise in our defective chip detection process. Consequently, it would be worth the team's effort to properly set the levels for these factors.

The proper settings are determined by the slope of lines depicted in Graph 1 or by examining the production run with the smallest signal-to-noise ratio. It seems logical then that production run 7 offers the best settings in our attempt to minimize the time it takes to detect defective chips. We refer to run 7 as our "experimental champion." Since factors E, F, and G are not significant, we can construct a "paper champion" which sets E, F, and G to 0. The complete prediction equation and predicted average response would be: $\hat{y} = 632.29 - 62.25(A) - 90.25(B) + 22.25(C) - 47.75(-A*B)$; predicted = 409.79 (coefficient for A is negative and we wanted to move predicted y in the minimum direction, so A was set to 1, B to 1, and C to -1). Furthermore, we could examine other interactions by weighting them to 0 and run a design using the S/N ratios. The power of the regression equation lies in its ability to predict our criterion response when we set the levels of our primary factors.

If our process average and standard deviation are fairly stable over time we may assume that the process is under control and can then assess the capability of the process through the use of metrics such as C_p or C_{pk} indices (Kiemele & Launsby, 1991). We decided to go with C_{pk} to take into account deviations from our desired target value of 600 msecs. We computed C_{pk} using the minimum of $((USL - \bar{Y})/3\sigma)$ or $((\bar{Y} - LSL)/3\sigma)$. Our minimum C_{pk} is .53. We would like to see C_{pk} values greater than 1.33. As our C_{pk} value goes up we would have fewer defects per million (dpm). We could also compute a loss-dollar value using Taguchi's loss function ($L = k(Y - T)^2$). The letter [k] represents our monetary constant. We could compute an estimated loss associated with any given Y-value and by multiplying average loss times by the total number of product produced we can estimate the total cost of our current state of quality (Kiemele & Launsby, 1991). It follows then, from our C_{pk} value that considerable changes in our defective-chip identification process are required in our attempt to improve quality.

Stage Three: Implement Feasible & Desirable Changes

Realistically there are three kinds of potential changes in a soft system: (1) changes in structure; (2) changes in procedures; and (3) changes in attitudes. Our simulation offers a mathematical model for making changes in structure and procedures. However, the hardest type of change to implement is in attitudes. Changes in structure and procedures may lead to unexpected and perhaps undesirable changes in attitudes of

the affected people in the soft system. Consequently, even after all the numerical analysis have been accomplished, it is important for our team members to define what is systemically desirable and culturally feasible. By systemically desirable we mean to say that those are changes which the model, for instance, suggest are important to make to improve quality. Whereas, culturally feasible changes refers to changes that the system can accept as meaningful and possible. To only go with systemically desirable changes, at the expense of what is culturally feasible, runs the risk of dire consequences.

Stage Four: Systemic Iteration

Once a system has cycled through the above three stages it is important to implement an iterative-continuous improvement process similar to the model we've described in this paper. This is absolutely critical since soft systems are dynamic. The mathematical computations described in this paper offer a systematic or methodical approach to improve, but there's more: for complex dynamic systems, it is necessary to be systemic, to be aware of what is needed in a dynamic and complex system. Without awareness of the whole system, the process can easily turn into a ritualistic adherence to procedures which no longer serve any function or purpose.

Conclusion

While our simulation centered on a chip manufacturing process, we believe the model described in this paper has a wide range of applicability for other soft systems. The same model, for instance, could be used to examine learning systems in the secondary or post-secondary learning environment (see Porter, 1991).

Note: The data used in this paper came from ChipSim, a computer software program which simulates a chip manufacturing process developed by the authors. The ideas and opinions expressed in this paper are those of the authors and do not necessarily reflect the view of the US Air Force Academy or any other agency of the Department of Defense.

References

- Archibald, K. A. (1989). A systems notebook . Unpublished manuscript.
- Checkland, P. (1981). Systems thinking, systems practice . New York: John Wiley & Sons.
- Kiemele, M., & Launsby, B. (1991). TQM tools and techniques: Measurement/statistical tools. In K.D. Lam, F.D. Watson, & S.R. Schmidt (Eds.), Total quality (pp. 9-1-9-34). Colorado Springs: Air Academy Press.
- Porter, D. B. (1991). Total quality education. In K.D. Lam, F.D. Watson, & S.R. Schmidt (Eds.), Total quality (pp. 11-1-11-29). Colorado Springs: Air Academy Press.
- Schmidt, S. R., & Launsby, R. G. (1989). Understanding industrial designed experiments. Colorado Springs: Air Academy Press.
- Senge, P. M. (1990). The fifth discipline . New York: Doubleday/Currency.



92-17651

Student Course Critiques:
Case Study in Total Quality EducationDavid B. Porter, DPhil
United States Air Force Academy

Abstract

This paper suggests the principles of Total Quality can be applied directly and effectively to higher education. The development and implementation of a comprehensive, theory-based instructor critique program at the USAF Academy is an example of one such application.

Total Quality and Education

Expanding the implicit boundaries around each organizational function, allows larger causal systems to emerge (Senge, 1990). Insight from such general models, in turn, provides a context for gradual but continuous improvement through the elimination of waste, improvement in the efficiency of separate activities, and greater synchronization of groups within the organization. Of equal importance is the development of trust and the displacement of interpersonal competition and animosity (Porter, 1992). At an organizational level, these issues, tensions and activities help define Total Quality (TQ). Aguayo (1991) suggests Total Quality should also be defined at the individual level: "people have to be happy for there to be quality" (p. xi). Deming's affirmation that "Quality is pride of workmanship" (Aguayo, 1991, p. xi) reiterates the positive impact of TQ at personal as well organizational levels. There is evidence Total Quality can enhance the effectiveness of individuals, organizations and even entire industries.

The tremendous increases in productivity and world wide competitiveness of the Japanese automotive and electronics industries over the past 40 years clearly demonstrate TQ's potential value. Consistently reducing variation in critical processes and eliminating activities which do not contribute to the quality of the product or service, result in increased productivity, decreased costs and prices, increased market shares and secure jobs for workers (Walton, 1986). There is little reason to believe Total Quality's success should be limited to a single culture or type of industry. Through such programs as the Malcom Baldrige National Quality Awards, a growing number of diverse American corporations achieving successes similar to those in Japan.

Nonetheless, the application of Total Quality to education has been rather limited. Marchese (1991) suggests that "the number of (institutions) committed to TQM on an institution-wide basis stands at two dozen... the number with deeper experience constitutes a mere handful" (p. 9). Similarly, Seymour and Collett (1991) found that even among those institutions adopting TQ, most initiatives were aimed at enhancing administrative efficiency rather than directly improving education. Contrary to these reports of academia's reluctance to accept Total Quality, it might be argued that, at their best, education and TQ are already intertwined in a relationship which is fundamentally symbiotic (Porter, 1991a). In fact, many of TQ's principles (although not specifically identified as such) underlie several successful academic programs (e.g., Alverno's commitment to instill lifelong skills and Harvard's Assessment Seminars). Perhaps, the Air Force Academy's development and implementation of a comprehensive instructor critique program is another example of a successful application of TQ principles to education.

What happened here?

The longest journey begins with a single step but the direction of that first step is often uncertain. Without a map and a compass, even the most well-intentioned quest can degenerate into hapless wandering. A conceptual model of education is a prerequisite. A general model focused on increasing the potential intellectual contributions of graduates emerged from informal faculty discussions (Porter, 1988; 1991a; 1991b). Students were seen to be willing workers rather than external "customers". Both Glasser (1990) and Marchese (1991) emphasize the importance of this perspective. Another key feature of this general model was the identification of three separate channels of student development (viz., knowledge, thinking, and affect). It was further hypothesized that graduates' potential contributions would be reflected by the multiplicative product of development along each channel. This suggests the greatest improvement in overall quality will result from increases in the channel which currently has the lowest absolute value (Porter, 1991a).

The next step was to use the model to assess current conditions. The Academy charges no tuition and offers excellent education opportunities, so there are always more qualified applicants than available appointments. As a result, the credentials of entering students are exceptional (e.g. average SAT scores of 1250, over 80% having graduated in the top 10% of their high school class, over 20% class presidents or vice presidents and over 80% having earned varsity letters (Foerster, 1990)). The belief among many of the faculty, however, was that cadets viewed education as merely something to be endured until graduation and flight school. These beliefs received support when a sample of 350 juniors completed Weinstein's "Learning Aptitudes and Study Skills Inventory". Cadets scored near national norms on 8 of the inventory's 10 scales, however, their scores on motivation to study placed them in the bottom quartile and their attitudes toward academics were at the 15th percentile. Earlier, Clover and Porter (1980) found the number of cadets dissatisfied with the academic program increased steadily with student tenure (from 5% for freshmen to over 25% for seniors). None of this suggests cadets performed poorly by objective national standards; it simply suggested their level of achievement was below their potential. No matter how much diverse knowledge our graduates had learned, nor how many esoteric computations they could perform, their apparently negative attitudes might limit their potential contributions as Air Force officers.

Initial assessment data generated a shared concern but not much agreement on how to correct the problem. Many potential approaches to this issue were considered and attempted in different academic departments. One of these was the development of a student critique to provide instructors with systematic feedback about the effect they had on students' knowledge, thinking and attitudes. Initial development of the critique within a single, 40-person department had several benefits: communication was easier, costs were lower, and objective evidence concerning validity, reliability and utility could be collected. All instructors in this department were invited to participate in the creation of the initial instrument. Instructors responded to three questions: "What do you do in the classroom that contributes to student's knowledge, thinking skills or attitudes?"; "If you were really successful, what would student's say about your class?" and; "What other factors (e.g., texts, tests, classmates, etc.) influence student development?" Their responses were used to create a common instrument which was administered to all students at the end of each semester. A "trusted agent" collected data

and tabulated results. Instructor's data were provided directly to them and not seen by supervisors or administrators. Two years of testing and development led to gradual improvement and increased confidence among department personnel.

One unique aspect of this critique was of particular value: the three separate channels of student development. A pedagogical theory developed by Broudy and elaborated by Hudak and Anderson (1984) suggests teachers place different relative emphases on students' knowledge, thinking skills and enjoyment (attitude). By creating an ipsative (zero sum) profile, teachers could be shown where their students saw them placing the greatest emphasis. This relative emphasis is both theoretically and computationally orthogonal to the teacher's level of perceived effectiveness (viz., the sum of their contributions to development on each channel). A common criticism of student critiques is that they are mere popularity contests which reward only teachers with a particular style. The specification of a style provided diagnostic information which clearly identified teachers' relative emphasis in the classroom.

Gradual refinement over 18 months produced an instrument which appeared to have merit. Several lessons were learned about the critique during its development. Instructor's ratings from students were compared with ratings they received from the department's senior faculty observers. Over three semesters, these correlations ranged from .57 to .74. Reliability for individual instructors from one semester to the next was found to vary between .73 and .85. In an extensive review of the literature on critiques, Cashin (1990) provides broad and consistent general support for instructor critiques. After reviewing the evidence, the 19 department heads agreed on the potential value of course critiques and chartered a committee with representatives from each academic department to develop a comprehensive program for the faculty.

The initial instrument, which has now been used for four consecutive semesters, contains 43 statements with which students indicated their relative agreement. The first 19 items contained specific classroom behaviors (e.g., my instructor: asked challenging questions, communicated effectively, listened to students, etc.). Six criterion statements (e.g., positive role model, thinking about material outside of class, working harder than expected, etc.). The remaining questions asked students the extent to which other factors (e.g., tests, texts, classmates, etc.) had influenced their enjoyment, thinking and subject knowledge. Students used a 10-point scale (0-9) with a neutral average response being 4.5. Using so many response categories is psychometrically unnecessary, but nonetheless useful. Shulman (1990) argues that it is precisely those students we often refer to as "the wrong kind" who most need quality teaching. The extended scale afforded those students who were most dissatisfied, extra influence on overall averages. This instrument was first administered in the Fall semester of 1989. Although a common instrument and feedback forms were used, each department determined their own internal uses of the instrument. "Trusted agents" consolidated data, conducted analyses and presented overall results. Faculty, divisional, year-level and course norms were provided as a part of every feedback package.

What's been learned about education at USAFA?

In the most general terms, we've found that an institution-wide course critique system is viable. Before the critique could have any impact, however, it had to be accepted and used by both students and faculty. In a recent survey of 959 students, over 90% agreed or strongly agreed with the

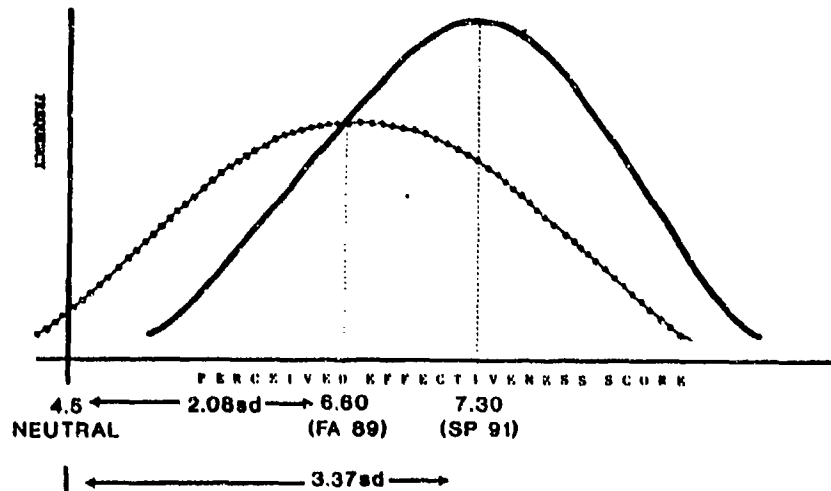
statement "I made a serious effort to complete this critique properly". Although the initial faculty response to survey questions concerning the critique's usefulness were only moderately positive, ratings improved markedly the following year. There is still considerable variety in the way critiques are used in the different departments, but all now provide instructors with complete individual critique feedback packages. As a common metric for evaluating one aspect of cadet's educational experience, the critique has facilitated exploration of a wide variety of research questions. Several of these are particularly noteworthy.

After graduation, cadets receive commissions and become Air Force officers. One of the qualities the Academy specifically seeks to develop is the ability "to work effectively and compassionately with others". For this reason, collaborative learning has become an increasingly attractive pedagogy. Several early studies with group learning showed remarkable improvements in student performance (viz., about one half standard deviation with a 50% reduction in variance between students). However, the question remained as to what actually caused these improvements. Although not conclusive, data from the course critique were relevant. In comparing ratings from the experimental sections (those using collaborative learning) with traditionally taught sections, the greatest differences were the perceived contributions of "classmates" to students' enjoyment ($t(52)=2.83$), knowledge ($t(52)=4.95$) and thinking ($t(51)=6.19$) (Porter, 1989). Similarly, when significant improvements in performance were recently noted in a core engineering course, critique data showed that cadets attributed much of the change from previous semesters to increased contributions of their classmates.

The teaching style data were also interesting. As suspected, faculty members who emphasized student enjoyment received general effectiveness ratings (viz., the average of six criteria statements) much higher ($\bar{x}=7.7$, $sd=.6$) than those who were seen by their students as emphasizing either knowledge ($\bar{x}=6.3$, $sd=.8$) or critical thinking ($\bar{x}=6.4$, $sd=1.4$). (Those with relatively equal emphasis had a mean rating of 7.1 and a standard deviation of .7.) Of particular interest, however, were the relative low ratings and high variance among teachers who were seen to place extra emphasis on critical thinking. This suggests that even if critical thinking is a desired quality for graduates, the direct approach may not be the best. Fortunately, indirect approaches such as collaborative learning, the inclusion of well-designed projects or the addition of provocative articles were all found to increase student ratings of a course's contribution to their critical thinking. Students also rated the critical thinking contributions of courses in their academic major almost a full point higher than they rated core course contributions. Additionally, it was found that teachers' contributions to critical thinking in particular improved significantly, the second time they taught a particular lesson.

Although many factors can affect student learning and development in a course, teachers were the largest single source of variance in the student attitudes reflected by their course ratings. The development and implementation of a theoretically based, comprehensive critique program is possible. Direct provision of feedback to faculty without mediation (i.e., implicit threat) can result in overall increases in the perceived effectiveness of instruction throughout the institution. The grand mean for the Perceived Effectiveness Score (an average of six criteria statements), which was comprised of six criterion items, increased significantly over the first four semesters of the critique's use. The following graph reflects the mean ratings and standard deviations of nearly 500 faculty members for the

CHANGE IN FACULTY PERCEIVED EFFECTIVENESS SCORES OVER FOUR SEMESTERS



first semester the critique was implemented (Fall of 1989) and two years later (Spring of 1991). These results look like quality, feel like quality and even taste like quality improvement; perhaps, they really are.

Note: Views expressed in this paper are those of the author and do not necessarily represent the views of any government agency.

References

- Aguayo, R. (1990). Dr. Deming: The American who taught the Japanese about Quality. New York: Carol Publishing Group.
- Cashin, W.E. (1990). Student ratings of teaching: A summary of the research. Management Newsletter 4(1). pp. 2-7.
- Clover, W.H. & Porter, D.B. (1980). Cadet Opinions of the best and worst aspects of the USAFA academic program. Psychology in the Department of Defense - Seventh Symposium. USAFA-TR 80-12. pp 105-112.
- Foerster, R.Y. (1990). Descriptive Characteristics and Comparisons for the Class of 1994. USAF Academy, CO: Office of institutional Research.
- Glasser, W. (1990, February). The Quality School. Phi Delta Kappan, 425-435.
- Hudak, M.A. & Anderson, D.E. (1984, October). Teaching style and student ratings. Teaching of Psychology, 11. pp. 177-178.
- Marchese, T. (1991, November). TQM Reaches the Academy. American Association of Higher Education Bulletin 44(3). pp. 3-9.
- 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, pp. 473-477.
- 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. 507-511.
- Porter, D.B. (1991a). Total Quality Education. in K.D. Lam, F.D. Watson & S.R. Schmidt (eds.) Total Quality: A textbook of strategic Quality leadership and planning. Colorado Springs, CO: Air Academy Press. pp 11-1 - 11-29.
- Porter, D.B. (1991b). A Perspective on College Learning. Journal of College Reading and Learning 24(1) pp. 1-15.
- Shulman, L.S. (1989, June). Toward a Pedagogy of Substance. American Association of Higher Education Bulletin. pp. 8-13.
- Seymour, D.T. & Collett, C. (1991). Total Quality Management in Higher Education: A Critical Assessment. Methuen, MA: Goal/QPC.
- Walton, M. (1986). The Deming Management Method. New York: Putnam Publishing.



Does the United States Air Force Academy's Training Philosophy Really Work?

Lieutenant Colonel Michael A. Rosebush, Ph.D.
USAF Academy

Abstract

This paper provides a summary review of the effectiveness of the United States Air Force Academy's "Academy Training Philosophy" (ATP), based upon all known studies. Results at both the U.S. Air Force and Coast Guard Academies indicated: (a) positive acceptance of both the model as a whole and with the individual ATP principles; (b) ATP behaviors correlated strongly with supervisory effectiveness ratings; and (c) military performance ratings of freshmen subordinates were significantly higher for those supervisors using ATP than those supervisors not using ATP.

Since 1984 the United States Air Force Academy (USAF) has used a set of principles for supervision and leadership known as the "Academy Training Philosophy" (ATP) [Rosebush, 1985], for the purpose of helping officers and cadets achieve a mutual respect relationship with their subordinates while maximizing performance. There are numerous supervisory behaviors endorsed in ATP, clustered into five main components: (1) setting clear expectations for the subordinate; (2) providing sufficient rehearsal opportunities for the subordinate to acquire the necessary skills for the job; (3) providing on-going feedback about the subordinate's performance; (4) providing appropriate consequences commensurate with the previous feedback; and (5) providing growth opportunities through establishment of strong self-esteem and realistic challenges.

The ATP has also been utilized at other Air Force settings (e.g., the Military Airlift Command and the Strategic Air Command), other Air Force commissioning programs (e.g., Officer Training School and the University of Pittsburgh Reserve Officer Training Corp), sister-service training centers (e.g., the U.S. Military Academy, the U.S. Coast Guard Academy, and the Citadel), and at civilian training centers (e.g., the North Carolina Highway Patrol Academy and the University of Minnesota School of Education). Although ATP continues to be used at numerous training institutions, a logical question becomes: "Does the Academy Training Philosophy really work?" This paper provides a review of the known findings regarding the effectiveness of ATP.

Receptivity of the ATP Model

One way to assess ATP's effectiveness is to ask the users what their subjective impressions of ATP are and whether the application of the model results in effective supervisory performance. The only known study of ATP receptivity occurred at the USAFA during the first summer in which the model was introduced (Rosebush & Bryant, 1984). Officers and cadets were surveyed

92-17652
■■■■■■■■■■

both immediately after they received initial ATP training and again after they had a chance to apply ATP's principles during Basic Cadet Training (an intensive military indoctrination program involving senior cadets supervising and training the incoming freshmen cadets). In the survey administered to 22 officers and 386 cadet cadre immediately after the initial ATP training, 93% of the respondents believed that ATP was a relevant approach to military indoctrination training, plus 97% believed "that ATP should be used by a supervisor to a subordinate". Similarly, the follow-up survey of the ATP trained officers and cadet cadre showed that 89% described themselves as successful in working with subordinates when applying ATP; only 4% responded that they were unsuccessful. Additionally, when asked "Should ATP be used again next Basic Cadet Training?", 73% said "yes" and 20% said "no". Finally, when asked "Should ATP be used during non-indoctrination times", 91% said "yes" and 4% said "no". These surveys indicated that officers and cadets who had received ATP training believed that the ATP model was relevant for military indoctrination training, beneficial to supervisor-subordinate relationships, and resulted in successful supervision when applied.

Stronger support for the effectiveness of ATP was gathered by USAFA's Office of Institutional Research, which annually asks all "basic cadets" (i.e., incoming freshmen) to describe "What was the most motivating experience for you during this BCT?" In the year previous to the introduction of ATP (i.e. the class of 1987), 11% of the basic cadets indicated that working with the cadet cadre was their most motivating experience (Rosebush & Bryant, 1984). However, when this same question was asked of the basic cadets during the summer in which ATP was first applied (i.e., the class of 1988), 21% of the basic cadets now responded that the cadet supervision was their most motivating experience. Similarly, when the class of 1987 was asked "What was the most demotivating experience for you during this BCT?", 41% responded "working with the cadet cadre" (this was also the most frequent negative response for that year). Dramatically, when this same question was asked of the class of 1988, only 3% of them indicated that the cadre was their most demotivating experience of BCT. While the positive receptivity toward the cadet cadre cannot necessarily be attributed to ATP (although the addition of ATP, for the most part, was the only difference from the previous year) there was nonetheless a rather remarkable change in perceptions toward cadet supervision in only a one year period.

Acceptance of the Individual ATP Concepts

Whereas the Rosebush and Bryant (1984) study provides evidence of support for the ATP model as a whole, a related question concerns "Do the individual concepts espoused in ATP really work?" In a three-year study of ATP beginning in 1986, 93 randomly selected cadets (all familiar and trained as supervisors in ATP's principles) wrote 848 behavioral items which they believed accurately described ATP-style supervision (Rosebush, 1989). Many of these items were redundant and therefore eliminated, producing a reduced list of 117 non-redundant items.

This reduced list was given to 140 randomly selected cadets (different from the first group of cadets, but also trained in ATP's principles), who were asked to (a) determine if each item really was an ATP principle, and (b) determine whether each item was one which "you have seen the really good cadet supervisors at USAFA do more of than the average supervisors, and that the average supervisors do more of than the poor supervisors". Of the 117 items, 44 of them (representing principles from each of ATP's five main components) received unanimous agreement as items which really represented ATP principles. Of these 44 remaining individual concepts, two of them (5%) were seen as effective in discriminating superior from average (and, average from inferior) supervisory performance by 90-100% of the cadet judges. Similarly, 24 of the individual concepts (55%) were rated as effective by 80-89% of the judges; 13 concepts (29%) were rated as effective by 70-79% of the judges; 4 individual concepts (9%) were rated as effective discriminators by 60-69% of the judges; only 1 item (2%) was rated as effective by 50-59% of the judges; and, none of the 44 concepts were seen as ineffective discriminators by a majority of the judges. This study seems to provide support that the individual concepts espoused in ATP are concepts which tend to be applied more frequently by the really good supervisors, as compared to the average or poor supervisors.

Relationship of ATP Concepts to Perceived Supervisory Effectiveness

If the Academy Training Philosophy really works, then one would expect to see a significant positive relationship between a supervisor's application of ATP's principles and the perceived effectiveness of the supervisor. An 86 item survey was given to 243 randomly selected cadet subordinates at USAFA to assess both the frequency of their cadet supervisors' application of ATP principles and the subordinates' perceptions about their supervisor (Rosebush, 1989). The 86 items were used to construct several ATP and outcome scales. The three outcome scales assessed the amount of respect for the supervisor, the amount of motivation subordinates feel toward achieving their full potential, and the amount of loyalty toward wanting to continue to work for their supervisor. The correlation between the 65 aggregated ATP items and the "respect" scale was .74. Similarly, the correlation between the combined ATP items was .77 with the "motivation" scale and .75 with the "loyalty" scale. All three correlations were significant at the $p < .001$ level.

Similarly, a 93 item survey was administered to 264 randomly selected USAFA cadet subordinates (different from any cadet group previously assessed) in order to determine the relationships between ATP's principles and perceived supervisory effectiveness (Rosebush, 1989). Forty-eight items (which had a combined Cronbach's alpha value of .97) were used to assess the cadet supervisor's frequency of applying ATP's principles, and were correlated with three different outcome scales used to measure the supervisor's perceived global effectiveness. The first global effectiveness scale was a two-item scale known as the "Supervisor Evaluation Questions" (which has a correlation of .83 between

the two items). The second and third supervisory effectiveness scales were the "job-in-general" and the "supervision" scales from Smith, Kendall, and Hulin's (1982) "Job Descriptive Index" (JDI). Simply put, these three scales were used to assess subordinates' perceptions regarding their supervisor's overall effectiveness. The correlation between the combined 48 ATP principles and the "Supervisor Evaluation Questions" was .80, while the correlation with the JDI's job-in-general scale was .71, and .65 with the JDI's supervision scale. All three correlations were significant at the $p < .001$ level. Certainly, correlations never prove causality, but there appears to be a very strong relationship between supervisors' application of ATP's principles and three global assessments of supervisory effectiveness.

Does the Application of ATP Improve Subordinates' Performance?

Much of the evidence in support of the Academy Training Philosophy reported thus far is predominantly subjective assessments by the supervisors and subordinates. A more difficult criterion to satisfy is whether the performance of subordinates whose supervisors apply ATP's principles is better than the performance of subordinates whose supervisors do not apply ATP's principles. The only known study of this type occurred when 457 USAFA cadet subordinates were administered a 48 item survey to assess the frequency in which their cadet supervisors were applying ATP's principles (Rosebush, 1989). USAFA's Military Performance Average (MPA) was used as an outcome measure of subordinates' military performance. The MPA is a combination of officer and cadet evaluations of any given cadet, and is computed on a rating scale much like an academic grade point average (i.e., scores ranging from a high of 4.00 to a low of 1.00). First, the cumulative MPA (i.e., the average of all MPA scores ever obtained) of each cadet in the study was obtained as a reference point for how strong a military performer each cadet had been prior to the semester being studied. Next, each subordinate rated his/her cadet supervisor on how frequently the 48 ATP principles were being applied. Finally, the subordinates' end-of-semester MPA rating was obtained. The hypothesis was that the subordinates of those supervisors who were applying the ATP would have a significantly higher MPA rating (after the subordinates' previous cumulative MPA ratings had been partialled out of the current semester's MPA ratings) than the subordinates of those supervisors who were not applying ATP. This test of the semipartial correlation, using only freshmen inputs, did prove to be statistically significant at the $p < .10$ level (which is the commonly accepted level for assessing statistical significance for semipartial correlations). In other words, the military performance of freshmen cadets did improve (above and beyond their previous cumulative performance) as a function of how frequently the supervisor was applying ATP's principles.

Effectiveness of ATP at the U.S. Coast Guard Academy

While ATP is applied at numerous training settings other than the USAFA, the only known published study on ATP's effectiveness in another setting occurred at the U.S. Coast Guard

Academy (USCGA) [Blake & Potter, 1992]. The USCGA developed a survey for assessing both the frequency of their cadet supervisors' application of ATP's principles, plus several outcome variables. Significant correlations (at the $p < .01$ level) were found between the supervisor's application of ATP and the supervisor's: (1) "goal emphasis", (2) "positive motivation" expressed, and (3) the absence of "harsh treatment". Significant ($p < .01$) third-order partial correlations (i.e. evaluating one factor while controlling for the effects of the other factors) were also obtained on each of the outcome factors with the ATP factors. Additionally, correlations at the $p < .01$ level (ranging from .20 to .56) were found between the ATP factors and such individual outcome items as "admiration and respect for cadet supervisors", "cadre were good teachers", "I felt treated like a 'real' person", "I would not have wanted to leave my unit", and "this experience was a time of personal growth for me". Once again, correlational results prove nothing. However, the USCGA's results are remarkably similar to the USAFA studies, showing strong relationships between ATP and numerous positive outcome domains. Finally, the USCGA administered the above mentioned survey both in the first summer in which ATP was applied at their service academy (1987) and again in 1990. Results indicated that after applying ATP for three years the cadet supervisors' scores had significantly increased ($p < .001$) in usage of ATP behaviors, and that "negative affect" toward supervisors had significantly decreased (Blake & Potter, 1992).

Summary

Studies conducted on ATP consistently show support for its principles, as assessed by subjective perceptual measures and by military performance, both at the USAFA and at the USCGA. Nonetheless, much more study of ATP is needed, which will provide further insight into ATP's effectiveness.

References

- Blake, R.J., & Potter, E.H. III. (1992). Novice leaders, novel behaviors, and strong culture: Promoting leadership change beyond the classroom. In K. Clark, M. Clark, & D. Campbell (Eds.), The impact of leadership. Colorado Springs, CO: Center for Creative Leadership.
- Rosebush, M.A. (1985). Applying the Academy Training Philosophy. (Available from the Department of Behavioral Sciences and Leadership, United States Air Force Academy, CO, 80840).
- Rosebush, M.A. (1989). A behavioral observation scale for subordinate appraisal of supervisory performance at the United States Air Force Academy. Dissertation Abstracts International.
- Rosebush, M.A., & Bryant, C. (1984). Implementation of the positive motivation model. United States Air Force Academy: Department of Behavioral Sciences and Leadership.



Cadets' Perceptions of Leadership Development Activities
at the United States Air Force Academy ¹

Second Lieutenant Derek A. West
Ohio State University

Lieutenant Colonel Benjamin L. Dilla, Ph.D.
USAF Academy

Abstract

The purpose of this research was to identify cadet perceptions of the effectiveness of a wide spectrum of leadership development activities at the United States Air Force Academy (USafa). Sixty cadet leaders in three categories (athletic, military, and extracurricular) were questioned using a computerized survey. The computer program gathered background information, performance data, and ratings of mandatory and optional programs. The overall USAFA experience is rated high in terms of its impact on leadership development, but there is room for improvement in many programs.

Leadership development is the main goal of the USAF Academy, as stated in our mission statement of graduating officers "with the knowledge, character and motivation essential to leadership." Yet some have suggested the Academy is not fulfilling this mission. General T.R. Milton, USAF, (1989) warns, "There is something wrong" (103). He feels the Academy is producing officers who lack the commitment to be leaders and career officers in the military. This recent scrutiny echoes the arguments of several years prior that careerism, occupationalism, and centralization had produced a state where "something vital is missing in Air Force leadership" (Benton, 1981).

Recently, the curriculum at the US Military Academy (West Point) has been criticized for not properly preparing future leaders (Zais, 1990). In response to these and other criticisms, West Point has implemented a new cadet leader development system (Palmer, 1990).

Bass' (1990) review of the literature on leadership education, training, and development points out several lessons which provide guidance for designing and evaluating leadership development programs. His general findings are that activities should be designed with both organizational requirements and individual knowledge, skill, and attitude needs in mind. Many studies have found that experiential or on-the-job training is favored over classroom activities. Bass (1990) also cites many studies that found leadership development activities should be action-oriented rather than theory-oriented, and the need for congruence between training environment and the environment for which the trainee is being prepared.

¹ This research was completed as an independent study project in 1991 when Lieutenant West was a senior at the USAF Academy, under the advisement of LtCol Dilla. The ideas presented in this paper do not necessarily reflect the views of the USAF Academy, the US Air Force, or the Dept of Defense.

In a study of junior Air Force aircraft maintenance officers (AMO's), Morabito and Dilla (1985) found the most important leadership development activities, as rated by the AMO's, were work-related experiences. Formal education, such as Professional Military Education (PME), were seen as much less important. They found perceptions of greater leadership development to be associated with experience as indicated by higher rank and prior military service. Their research also supports Bass' (1990) conclusions in that they found experience should be specifically related to the organizational environment in order to be perceived as most effective (Morabito & Dilla, 1985).

What is the state of affairs at the USAF Academy? How effective is current leadership development perceived to be, both in an overall sense and with respect to specific programs? What activities do cadets view as most supportive of the goal of leadership development? To provide preliminary answers to these questions and to provide direction for future research, this study examined the perceptions of cadet leaders regarding the impact of a variety of USAFA programs on their leadership development.

Method

Subjects

The subjects of this study were 60 first class (senior) cadets at USAFA. Three groups of 20 subjects were randomly selected from the available pool of cadet leaders in three categories: (1) military--group and squadron commanders; (2) athletic--team captains; and (3) extracurricular--Cadets-in-Charge (CIC's) of USAFA clubs.

Apparatus

This research used a survey administered via computer disk on cadets' personal computers. This emerging technology was key to the success of the study and the accuracy of the results as we found cadets to be fascinated by the format of the survey. Audio and visual effects captured cadets' interest and yielded many positive comments--not the typical reaction to surveys at USAFA.

Use of the computer disk helped ensure data quality by not accepting invalid data and by eliminating the need to transcribe data; this also made the analysis phase faster and easier.

Procedure

Computer disks containing the survey program were sent to each subject through the cadet distribution system. Each disk was accompanied by a cover letter explaining the purpose of the study and how to start the program.

In the program itself, the subject was welcomed and again told the purpose of the research. Definitions of leadership and leadership development were presented for a common frame of reference. Then subject background information was gathered, including gender, academic major, grade point average (GPA), and military performance average (MPA).

The program pre-designated 21 broadly defined "leadership development" programs which are mandatory for all cadets, including such things as the core curriculum, fourth class system, weekend training events, and the cadet honor code. Cadets were asked to designate any of 39 other leadership development

programs in which they participated voluntarily, such as flying programs instructor, special summer programs, or other leadership roles not required of all cadets.

Next, cadets were presented with each of the programs in which they had participated, in random order, and asked to rate them for the impact on their leadership development using a seven-point scale from "Very Negative" to "Very Positive" with four indicating "Neutral." Finally, they were asked to rate the impact of their overall USAFA leadership development experience on the same seven-point scale.

Results

Data from the computer disks was aggregated and analyzed using the SPSS/PC+ statistical package for IBM PC's. With regard to background information, subjects' ages ranged from 20-25 years old. There were 18.3% females in the sample compared to about 12% females in the cadet wing. The average GPA for these cadet leaders was 3.18 (on a four-point scale) and MPA was 3.35. Thus, the "average" cadet in this sample was on the Dean's, Commandant's, and Superintendent's lists for outstanding performance-- a measure of success from the institution's perspective.

These cadet leaders were spread out among 22 different academic majors with the three most common being Behavioral Science--Human Factors (13), History (8), and Political Science (5).

Table 1 summarizes the ratings of programs for their impact on cadets' leadership development. Only those programs rated by at least one-third of the sample (20 cadets) were included. This cut-off ensures a reasonably stable average while recognizing there were three distinct groups, probably with differing experiences, who rated the programs. The programs are rank-ordered in Table 1 according to mean ratings, from most positive to most negative impact. Also included is the standard deviation of their ratings and the number of cadet leaders who rated each program. The rating of the overall USAFA leadership development experience is also included in Table 1.

Results show the overall USAFA leadership development experience is rated positively (mean of 5.9 on a seven-point scale). The activities which earned the highest ratings were the current leadership roles of these cadets-- Club CIC, Team Captain of a Sport, and Squadron Supervisor. (Group and Wing Staffs had equally high average ratings but were rated by insufficient numbers of cadets to be included in this tabulation.) Summer training program leadership positions (Survival, Evasion, Resistance and Escape [SERE] and First and Second Basic Cadet Training [BCT] cadre) also received some of the highest marks. Intercollegiate athletics, club activities, and church/religious activities also received high marks (equal to or above the overall USAFA rating).

Only one formal leadership education program received an average rating higher than the overall USAFA experience--the Commanders' Leadership Enrichment Seminar (CLES). This is a two-day workshop program for military leaders (wing, group, and squadron commanders).

The lowest rated programs included the cadet MPA system (which assigns military ratings to cadets twice a year), the Leader Attributes Survey (a cadet rating and feedback program), the cadet disciplinary system, morning and noon meal marching formations, and academic core courses as a group. Besides being the lowest rated items in a relative sense, the average ratings for these programs were at or below the "neutral" point on the rating scale.

Table 1
Ratings of Cadet Leadership Development Activities

<u>Program</u>	<u>Mean</u>	<u>Std Dev</u>	<u># of Raters</u>
Cadet-in-Charge of Any Club	6.68	0.568	22
Team Captain of an Intercollegiate Sport	6.64	0.790	22
SERE Cadre	6.36	1.026	28
Squadron Supervisor (Sq Cmdr, Element Ldr)	6.27	0.992	52
Second BCT Cadre	6.24	0.689	29
Intercollegiate Athletic Participation	6.22	0.929	36
First BCT Cadre	6.22	0.892	27
Commanders Leadership Enrichment Seminar	6.17	0.834	23
Church/Religious Activities and Clubs	5.94	1.029	33
Overall USAFA Experience	5.90	0.817	60
Extracurricular Clubs	5.79	1.382	24
Fourth Class System	5.67	1.374	60
Training of the Fourthclassmen	5.63	1.390	60
Basic Cadet Training	5.58	1.265	60
Squadron Staff	5.57	1.118	49
SERE as a Student	5.57	1.226	60
Beh Sci 310 (Academic Crse on Leadership)	5.47	1.120	49
Freefall Parachuting Program	5.42	1.474	26
Major's Courses	5.40	1.108	60
T-41 or Navigator Training	5.38	1.213	40
Operation Air Force	5.37	1.128	59
Coach of Any Intramural Sport	5.36	0.848	22
Intramural Sport Participation	5.13	1.209	47
Honor Code and Instruction	5.12	1.223	60
Squadron Sponsor Trips	5.05	0.964	60
Academy Training Philosophy (ATP)	4.97	1.288	60
CONUS Summer Program	4.91	0.996	54
Professional Military Studies	4.80	1.054	60
Academic Group Projects (e.g., Engr 410)	4.70	1.253	60
Intramural Sport Referee	4.65	1.531	20
Project Warrior	4.62	0.940	60
Soaring Program	4.47	1.419	51
Professional Military Trng (M-5 classes)	4.43	1.267	60
Supervisor Feedback (Forms 6 & 76)	4.42	1.253	60
Weekend Training (SAMI, Parade, IRI)	4.42	1.369	60
Cadet in Charge of Quarters (CCQ)	4.37	1.149	60
Squadron Duty Officer (SDO)	4.23	0.689	60
Academic Class Section Marcher	4.12	0.963	52
Academic Core Courses	4.07	1.148	60
Morning and Noon Meal Formations	3.90	1.285	60
Cadet Disciplinary System (Fms 10, Tours)	3.85	1.376	60
Leader Attributes Survey (Top/Bottom Four)	3.63	1.687	60
Military Perf Average Rating System	3.25	1.514	60

Based on previous research and some hypotheses about cadet behavior not discussed in this paper, responses were analyzed for differences across the three different categories of cadet leaders. Only one significant result was found. Consistent with predictions, athletic leaders were found to have participated in a smaller number of leadership development programs than cadets in the military/extracurricular categories (means of 35 versus 37 programs, respectively). The practical significance of these results will be addressed next.

Discussion

Although no statement of causality can be firmly stated based on these results, a plausible explanation for the difference in program participation is that athletic leaders simply do not have the time to participate in more development programs. A time management analysis or longitudinal study of these groups might provide more insight into this difference.

The preponderance of actual leadership role experiences among the highest rated programs is consistent with the findings of Morabito and Dilla (1985) for Air Force officers in the aircraft maintenance career field. The presence of the fourth class system, from both follower and leader perspectives, among the higher ratings speaks well for this aspect of Academy life. At the same time, however, it is curious that the Academy Training Philosophy (ATP), the guideline governing this system, is rated considerably lower (just under 5.0 for ATP compared to 5.7 for the fourth class system and 5.6 for training fourthclassmen).

Further research should explore the reasons for the results found here using greater numbers and a wider sampling of cadets. In particular, the negative ratings of some key institutional programs need to be understood. Some of the activities grouped together here (e.g., squadron supervisor) could be broken out to investigate what may be important differences between positions (Squadron Commander vs. Flight Commander vs. Element Leader). Continued assessment of how our programs and activities contribute to the goal of leadership development, and how they could be improved, is a must.

References

- Bass, B.M. Bass and Stogdill's handbook of leadership (3rd ed.). New York: The Free Press.
- Benton, J.C. (1981). Promoting leadership in the Air Force's management environment. Student report, Air University, Maxwell AFB, AL.
- Milton, T.R. (1989, March). The commitment gap. Air Force Magazine, p. 103.
- Morabito, M.A., & Dilla, B.L. (1985). Leadership development of USAF Aircraft maintenance officers. Proceedings of the 27th Annual Meeting of the Military Testing Association, 118-123.
- Palmer, D.R. (1990). Cadet leader development system. USMA Circular 101-1, West Point, NY: United States Military Academy.
- Zais, M.M. (1990, March). West Point: Sword making or swordsmanship? Armed Forces Journal International, 57-62.



The Quintessence of Quality

David B. Porter, DPhil
United States Air Force Academy

92-17654



Abstract

Total Quality's structure accounts for its success. Certain aspects of Total Quality are essential for its success. The increased knowledge and understanding which comes from a tension between systemic reflection and systematic measurement and analysis is one such essence. Another essence is the development of trust throughout the organization. Such trust requires both integrity and positive regard. Even trust and understanding together, however, are not sufficient to insure the success of Total Quality. The final essence, or quintessence, of quality is synchrony, which involves doing just the right things, in just the right way, and at just the right time.

There are a bewildering array of tools, techniques, and programs which claim to be essential parts of Total Quality (TQ). The ubiquity of such claims reflects the tremendous success generated by the original Total Quality concepts developed by W. Edwards Deming and others. There is, however, a danger that the truly essential aspects of total quality, which most directly led to success in the Japanese electronics and automobile industries over the last four decades, will be obscured by this plethora of pretenders. In one sense, total quality's success has become its own worst enemy. In some organizations, individuals feel obliged to take any idea and wrap it in TQ verbiage before presenting it. A common language can enhance communication but common words with diverse meanings and implications create confusion. Total Quality involves more than new labels for old processes or the discovery of new methods for achieving old objectives. Nearly seventy percent of the quality initiatives in the U.S. are failing. It's not the case that these companies do not accept Total Quality's general goals of focusing on customer satisfaction and continuous improvement; these sound like pretty good things to even the most traditional managers. What is often missing are the theoretical aspects of Total Quality which provide the medium through which Total Quality transformations occur. Peter Senge suggests many of the failures result from managers trying to implement techniques to achieve Total Quality goals without first adopting a true "quality philosophy" (Welter, 1991). Superficial changes by managerial dilettantes who pick and choose aspects of quality to enhance their image are likely to do substantial damage to their organizations and harm to their employees.

Total quality is not just another program or management fad. The paradigm shift total quality requires can lead to the discovery of new meaning and purpose in life as well as work. One of the criticisms sometimes leveled against total quality is that it is "almost a religion". In fact, the shift to a TQ perspective can be quite similar to a religious conversion. This is not because of the particular answers TQ provides, but because the questions TQ poses deal with issues which involve ultimate human concerns (Christian, 1973). Unlike traditional religions, Total Quality does not require irrevocable commitment to a particular dogma or set of beliefs. As Alan Watts (1957) suggests, such commitment "is not only intellectual suicide; it is positive unfaith because it closes the mind to any new vision of the world. Faith is above all openness—an act of trust in the unknown."

Nonetheless, Total Quality does require commitment. It requires commitment to a coherent philosophy which underlies many diverse processes and techniques. Just like language (Chomsky, 1965), Total Quality has both surface features and a deep structure. It is the deep structure which accounts for TQ's success and differentiates it from many of the more recent and superficial pretenders. One of Deming's favorite rebuffs to managers frustrated by the effort and integrity Total Quality demands is to remind them that TQ is not "instant pudding" (Walton, 1986). In contrast, Crosby's (1984) promise that quality can be both free and certain has great appeal to managers who prefer the "look" of quality to its ardor. Others have amplified the confusion by claiming that quality is basically eclectic and even asserting that the freedom to pick and choose features from different programs is itself an essential aspect of quality (e.g., Coate, 1990). Successful Total Quality transformations require profound understanding and exceptional integrity; anything less is simply "tampering" (Aguayo, 1990).

This is not the first time mankind has faced the problem of discriminating between substance and appearance. Ancient Greek philosophers used the term "ousia" or essence to distinguish the permanent from the variable and phenomenal. The four original essences were: fire, air, water and earth. Similarly, there are certain essentials which are necessary for Total Quality to work. One of these involves the analytic and deductive processes which work together to enhance understanding. The tension between the systemic and the systematic provide the power and the drive which allow continuous improvement and simultaneously enhance understanding of critical processes. Integrity and positive regard provide the guidance to keep transformations moving toward true Total Quality. The confluence of these two attributes creates trust throughout the organization which guides and sustains the momentum created by increased knowledge. Perhaps it would be useful to explain these first two essences (understanding and trust) before proposing a third.

Increasing Understanding (Acquiring "Profound Knowledge")

The systemic involves the development of a "systems perspective". Peter Senge (1990) defines systems thinking as "a conceptual framework, a body of knowledge and tools... to make the full patterns clearer, and to help us see how to change them effectively" (1990, p. 7). We study systems by creating abstract representations of critical causal relations which underlie organizational processes and activities (Cohen & Cohen, 1983). Argyris and Shon (1978) argue that although each of us has a "theory" which guides our actions, the theories we explain to others are actually quite different. Presented diagrammatically or mathematically, systemic beliefs become manifest in models and equations. Such systemic models are expansive and inclusive; they relate to the biggest of pictures and provide a context all ongoing activities.

Systematic assessment provides a necessary complement to systemic vision. Total quality managers must become applied scientists (Porter, 1991); once they have clearly exposed their beliefs and assumptions about important causal relationships, they must actively seek to disprove themselves by systematically and objectively observing actual processes. Contrary to their natural inclinations, they must seek data which might contradict their assumptions and a priori theories. What we think we "know" interferes with

our objectivity. The scientific method (Mook, 1982) as well as the tools and techniques of total quality were developed to identify and minimize such distortion through systematic measurement and analysis.

The concepts of evidence and proof are also relevant to the development of "understanding" or "profound knowledge". In general, science operates on a principle of indirect rather than direct proof. It is virtually impossible to unequivocally "prove" that a particular perspective is ultimately "the truth". Rather, the strategy is to collect evidence and see the extent to which plausible alternatives are likely to be correct. Social scientists in particular realize that there are no perfect experiments; they learn to continually refine and improve their models of human behavior as new evidence becomes available and look for convergent evidence from many different sources. Unfortunately without a systematic approach to collecting and analyzing data, most humans fall into the trap of relying on information which is convenient (i.e., individual anecdotes) rather than that which is valid or reliable.

Creating Trust (Turning Organizations into Institutions)

The second essence is trust, and as previously suggested, the development of trust depends on both integrity and positive regard. Total quality integrity requires honesty, humility and a willingness to interact with the organization as an institution rather than as a private indulgence (De Pree, 1989). An institution provides a common cause to which both workers and managers can commit themselves. In Total Quality organizations, individuals come to see their work much more as service to an institution than as a mere occupation (c.f., Moskos, 1977) (Perhaps Deming's use of the verb "institute" in several of his 14 points reflects a similar recognition of the necessity to create true "institutions".) Equity theory (Adams, 1965) suggests that individuals are likely to put forth effort in proportion to the value they receive relative to those around them. Managers' high visibility often provides a superordinate reference by which workers adjust both the quantity and quality of their inputs. Managers set the standard for commitment to the institution. One glaring difference in Japanese organizations committed to quality and traditional American corporations is the precipitous organizational height of most American corporations. It is not unusual for American CEOs to be paid hundreds of times more in salaries and bonuses than their workers. In Japan, the average compensation differential from bottom to top of the corporate hierarchy is about five and bonuses are often shared proportionately throughout the organization.

Trust requires more than integrity, however, it also depends on positive regard. Carl Rogers (1961) suggested unconditional positive regard was necessary for human health and development. Although the "unconditional" aspect of Rogers' prescription may not fit well with the traditional orientation of many contemporary corporations, the importance of positive supervisory expectations is well supported by both theory and research. McGregor (1960) contrasted managers with negative expectations of subordinates (Theory X) and those with positive expectations (Theory Y). Rosenthal (1968) studied the effects of supervisor's expectancies on the subsequent performance of their subordinates and found consistently strong positive relations between the two. Glasser (1990) used orientation toward students as one of the

principle ways to distinguish between traditional boss-teachers and the lead-teachers necessary for quality education.

Synchrony, the Quintessence of Quality

Although the ancient Greeks agreed all four physical essences were necessary, concern soon emerged that alone, these four were not sufficient. Likewise, understanding and trust are necessary but may not be sufficient to insure Total Quality. The Greeks solved the problem by creating a fifth and higher essence, the "quintessence". For them, the quintessence was "ether" from which all heavenly bodies were composed and to which those things that could not be explained by the other essences were relegated. The quintessence of quality is synchrony which reflects the timely interaction of the other essences and their components. Synchrony cannot replace the other essences; they both remain necessary. Rather synchrony defines the nature and timing of the relationships between the other factors (systemic vision, systematic inquiry, integrity and positive regard).

In contrast to many contemporary approaches to management and leadership, Total Quality stresses the importance of harmony and facilitation rather than coercion, pressure and force (Welter, 1991). "Just-in-time" training and production and "inventory-as-required" (Walton, 1986) are examples of how synchrony can enhance operational effectiveness. Synchrony involves doing the right things, in the right way, at the right time. Accomplishing this requires both a clear understanding of essential processes and a high level of trust throughout the organization. What is "right" is contingent on a myriad of other factors but those who try to force the action from "on high" only diminish quality in the long run. Crosby's "zero defects" and "do it right the first time" are the antithesis of synchrony. His personal anecdote (1984) about how his commitment to personal weight control prevented him from taking even a sip of champagne at his daughter's wedding raises a related issue. Such obsessive concern for absolute perfection may well alienate rather than impress subordinates. Synchrony requires that excess and ostentation be avoided. Even the essentials of quality can be stressed to such an extent they create their own opposites. For example, systems can expand to the point of becoming delusional and grandiose. In the extreme, systematicity becomes myopic obsession or anal compulsion. Even integrity can be exaggerated to the point of dysfunction and excessive positive regard can turn a leader into a soppy sycophant.

Synchronous leadership is quintessential to Total Quality. Deming's seventh point is "institute leadership" and several other points clearly pertain to leadership (e.g., create constancy of purpose, adopt the new philosophy, drive out fear, etc.). Glasser (1990) makes a clear distinction between traditional "boss-managers" and the individuals needed to lead Total Quality organizations (viz. "lead-managers"). Bosses set tasks and standards for subordinates, tell workers how to do things, inspect work, and when workers resist, use coercion to achieve compliance. In contrast, leaders engage workers in dialogues, model expectations, ask workers to inspect their own work, and nurture the growth of positive work environs. This approach to leadership is not new either. The following is a poem by a contemporary of the ancient Greeks:

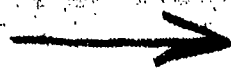
Leaders are best when people barely know they exist.
Not so good when people obey and acclaim them,
Worse when they are despised.
Fail to honor people and they will not honor you.
But of good leaders who talk little,
When the work is done and the aims fulfilled,
Their people will say 'We did this ourselves'.

The Chinese philosopher, Lao-Tse (565 B.C.) clearly has captured the synchrony which is the quintessence of quality.

Note: Views expressed in this paper are those of the author and do not necessarily represent the views of the U. S. Air Force nor any other government agency.

References

- Adams, J.S. (1965). Inequity in social exchange. In L. Berkowitz (ed.), Advances in experimental social psychology, vol. 2. New York: Academic Press.
- Aguayo, R. (1990). Dr. Deming; the American who taught the Japanese about quality. New York: Carol Publishing Group.
- Argyris, C. & Shon, D. (1978). Organizational Learning: A theory of action perspective. Reading, MA: Addison-Wesley.
- Chomsky, N. (1965). Aspects of the Theory of Syntax. Cambridge, MA: MIT Press.
- Christian, J.L. (1973). Philosophy: an introduction to the art of wondering. San Francisco: Rinehart Press.
- Coate, E. (1991) Implementing Total Quality at Oregon State University.
- Cohen, J. & Cohen, P. (1983). Applied Multiple Regression / Correlation Analysis for the Behavioral Sciences. Hillsdale, NJ: Erlbaum.
- Crosby, P.B. (1984). Quality without tears: the art of hassle-free management. New York: New American Library.
- De Pree, M. (1989). Leadership is an Art. New York: Doubleday.
- Glasser, W. (1990). The Quality School: Managing students without coercion. New York: Harper & Row.
- McGregor, D.M. (1960). The human side of enterprise. New York: McGraw-Hill.
- Mook, D.G. (1982). Psychological Research: strategy and tactics. New York: Harper & Row.
- Moskos, C.C. (1977). From institution to occupation: trends in the military organization. Armed Forces & Society 4(1). pp. 41-49.
- Porter, D.B. (1991). Total Quality Education. In K.D. Lam, F.D. Watson & S.R. Schmidt, (eds.), Total Quality: A textbook of strategic quality leadership and planning. Colorado Springs, CO: Air Academy Press.
- Rogers, C. (1961). On Becoming a Person. Boston: Houghton-Mifflin.
- Rosenthal, R. (1968). Pygmalion in the Classroom. New York: Holt, Rhinehart & Winston.
- Senge, P. (1990). The Fifth Discipline: the art & practice of the learning organization. New York: Doubleday.
- Walton, M. (1986). The Deming Management Method. New York: Putnam Publishing Group.
- Watts, A. (1957). The Way of Zen. New York: Pantheon Books.
- Welter, T.R. (1991). A winning team begins with you. Industry Week (May 6). pp. 35-42.



AD-P006 922



**Systems Science in Support of
Total Quality Management¹**

92-17655



Major Larry B. Rainey, Ph.D.
Air Force Space Command
Colonel Charles J. Yoos II, D.B.A.
Air Force Academy

Abstract

Systems science provides additional tools to augment the TQM "tool kit." This relatively new science and its basic four foci are introduced. The Conant-Ashby Theorem, from system science, is addressed and application is made through the organizational system model and the operational system model.

Systems science has much to offer Total Quality Management (TQM). This paper bridges the gap between these two subjects to illustrate specific means by which systems science can begin to augment TQM. This construction is based upon the consideration of two topics. First, a brief introduction to systems science is provided. Second, systems science is applied through the concept of the Conant-Ashby Theorem.

Introduction to Systems Science

We know what classical science is, but what is systems science? An introduction to this subject can start with understanding the distinction between thingness and systemness. For review, classical science, in general, concentrates on the differences between things. Specifically, it studies things themselves. Therefore, the result is a disciplined classification of the study of various categories of things. In contrast, systems science not only transcends the boundaries of the classical disciplines but it is also cross-disciplinary. As such, it "focuses on the study of systemhood properties of systems" (Klir 1991). Taking the above ideas into account, a formal definition can now be provided. "Systems science is a science whose domain of inquiry consists of those properties of systems and associated problems that emanate from the general notion of systemhood." (Klir 1991).

What are the characteristics and thrusts of systems science? Like classical science, it "contains a body of knowledge regarding its domain, a methodology for acquisition of new knowledge and for dealing with relevant problems within the domain, and a metamethodology, by which methods and their relationship to problems are characterized and critically examined" (Klir 1991). With this construct, it examines the ways things are organized, it studies operational relationships and its significance increases with growing complexity of systems (Klir 1991). Therefore, systems science "provides a paradigms for sound management of complex organizations" (Schoderbek 1975).

There are four foci for the systemic investigation of organizations. First, the systems researcher sees the organization as a system that is open, e.g., exchanges energy (information

¹ This paper represents the views of the authors and does not necessarily reflect the official views of the Air Force Space Command, the Air Force Academy or the Department of the Air Force.

and other resources) with its environment. Therefore, his/her world is made up of organic-open systems. In this context, two processes are of utmost importance: growth and control. "Growth is a necessary condition for the survival of any system; at the same time, control (the ability of a system to sustain a rate of growth in keeping with its capacity and the environment's tolerances) is a necessary condition for balanced growth." (Schoderbek 1975).

Second, systems science is holistic in its investigation of the organic system. This does not suggest that systems scientists study everything about everything. In operational terms, the holistic approach does not suggest that the manager (process owner) should be concerned about everything within his/her department or company. Rather, "holism implies that enough thought will be given to determining the critical variables influencing the growth and control patterns of an organization as well as to establishing ways of monitoring the critical parameters in the organization-environment interface" (Schoderbek 1975). In the above example, this approach requires the individual to go at least one step beyond what is considered satisfactory. (Schoderbek 1975).

Third, the investigation of organizations, within a holistic approach and under constantly changing conditions, is facilitated by modeling and simulation. This process begins with a gross conceptualization of the system and ends up with a refined model of the abstracted reality. (Schoderbek 1975).

Finally, the last foci pertains to the outcome of applying the systems science paradigm. The result should be an understanding of the situation under study as it relates to the rest of the organization and its environment. Having achieved this new vantage, quantification techniques (e.g., simulation) can be exercised to develop possible courses of action for the manager's (process owner's) consideration. (Schoderbek 1975).

Applications of Systems Science

Understanding and controlling processes is a major focus of attention for TQM. Beer (1979) has stated that systems (processes) are observer dependent as to their nature, purpose and boundaries. Therefore, there is a tool required to provide a common understanding of the organization as it strives to achieve its vision, mission and goals. In addition, a tool is required to provide an accurate description of processes as they operate within the organization. Such tools would provide a means to technically evaluate and analyze problems within the organizations as they arise, forecast the performance of the organization and assist in continuous improvement.

Conant-Ashby Theorem

If systems science is invoked to augment TQM, additional concepts and tools are made available to contribute to the TQM "tool kit." One of these is the Conant-Ashby Theorem (Ashby 1970). In its pure form, this theorem states that "every good regulator of a system must be a model of that system." Rainey (1991) has modified this theorem to state, "every good regulator of a system 'must contain' a model of that system." Within a management context, Rainey (1991) has defined the regulator to be the manager (process owner) responsible for the system (process) under study. Therefore, to be a "good regulator", the manager (process owner) must have a model of the system (process) he/she is responsible for managing. The Conant-Ashby Theorem, in its application through systems analysis, leads to the application of two models (tools): the organizational system model and the operational system model.

Organizational System Model

The organizational system model is not an organization chart, for this chart only illustrates who reports to whom. Rather, an organizational system model is a gestalt or synergistic representation of how the various operations of a corporate organization function together to achieve the stated vision, mission and goals of the organization. It provides management with a coherent and single picture of the total organization, identifying logical relationships and the channels of communication through which flows the life blood of the organization-information and not just data. The pictorial representation has an important practical benefit. By providing executive managers with a synoptic view of their system, they can improve their understanding of organizational behavior. Without such a focus, each manager may be using a slightly different system model (mental perception) to gain such an understanding. The result is a lack of congruence in that each decision maker may be, in effect, controlling a different organization. The model provides a corporate focus for command and control (Pearson 1985).

Schmidt (1990) has addressed the need for a systems view of the organization. He stated that TQM also addresses cross-functional operations and goals for continuous improvement. Quality is not the sole function of one department or functional area; it is a cross-functional goal. The same idea applies to achieving improved efficiency or providing customer satisfaction. The organizational culture must stress a system's view of the organization.

There are at least two candidates (Beer 1979, Gharajedaghi 1985) for the organizational system model. We would elect the use of Beer's (1979) Viable System Model (VSM). It can be applied to each major level of the corporate organization to test whether each level is performing adequately in context of each of the five major functions addressed by the VSM. Another author (Leonard (no date)) has reached the same conclusion and has documented the general application to TQM.

Operational System Model

This model is the plant or process model as found in control engineering (Hostetter 1982). In other words, it represents the relationships associated with the operation under study. Four components are associated with the operational system model: the controls, the controlled variables, the state variables, and the plant or process model itself. The controls are the inputs to the plant model, the controlled variables are the outputs of the plant model, and the state variables describe the fundamental internal variables of the plant or process. The plant or process model consists of two submodels: the structural model and the parametric model (Beer 1966).

The structural submodel is the "equations of motion" of the operation under study. Ideally, this submodel would represent the operation via a series of mathematical equations. If this is not possible, an attempt should be made to characterize the operation via a simulation language (e.g., SLAM or Simscript).

The parametric submodel is the quantification of the structural submodel. In this submodel, the developer provides "actual quantities that are descriptive of the world situation" (Beer 1966). This may involve specific probabilities, coefficients and other numerical data that characterize the actual operations. Ideally, the end product of the operational system model will be a computer simulation.

How does one develop the operational system model? One of the tenets of TQM is continuous process improvement. Several strategies can be implemented in this regard. An obvious candidate is the Deming Cycle. The first step in the Deming Cycle is, recognize the opportunity. "To recognize the opportunity, you must first define your process." (Schmidt 1990). We propose the following five steps ((Rainey 1991) to define the process with respect to the OSM.

The first step is to develop a workflow diagram, a visual representation of the flow of work from start to finish. The benefit of such a diagram is that it provides a synergistic understanding of the actual day-to-day operations under investigation. Beer (1979) states that

systems are observer dependent as to the definition of nature, purpose and boundaries. Therefore, staff coordination is required for the synergistic perspective.

The second step is the development of the simulation network diagram. At this point, it is necessary to determine the best type of simulation language (e.g., discrete event, analog, etc.) to meet the developer's requirements. Third, from the simulation network diagram the simulation software code follows immediately. The fourth and fifth steps are the customary simulation exercises of verification and validation. Verification checks the internal integrity of model relationships. Validation assures that the simulation replicates the real world, by comparing simulation output with the actual world results being modeled.

Benefits of Applying the Conant-Ashby Theorem

The primary problem associated with system (process) control is perception. This was confirmed by Beer's (1979) observation that systems (processes) are observer dependent as to their nature, purpose and boundaries. For the corporate organization, we stated that the primary benefit of the organizational system model was that it "provides management with a coherent and single picture of the total organization identifying logical relationships and channels of communication..."

There are five benefits derived from developing the operational system model. First, it provides a vehicle to determine the controls, state variables and controlled variables for the process under study. Second, the monitoring of the controlled variables facilitates management by exception. Third, once the controlled variables are determined, appropriate statistical control routines (e.g., Kalman Filters) can be applied for short term forecasting. Fourth, this model can be used as a standard for comparison to actual operations. When differences exist between desired and actual operations, managers have a tool for troubleshooting. They can evaluate proposed solutions, before implementing them, to assess their impact. Finally, this model can be used for long range planning. If a process requires modifications, managers (process owners) can determine which proposal is most cost-effective. As such, it can be used to test which policies are preferred with respect to which courses of action to follow for the future.

Summary

Systems science has much to contribute to TQM. This contribution is realized through the Conant-Ashby Theorem. Application of this theorem is in two parts: the organizational system model and the operational system model. Both of these have their own specific benefits but the general benefit is to provide a common focus of attention for the corporate organization and process under investigation. The practical benefits of the ideas addressed in this paper will only be realized once those involved in TQM implementation apply them to real world problems.

REFERENCES

- Ashby, W.R. and R. Conant. (1970). "Every good regulator of a system must be a model of that system." International Journal of Systems Science, 1:89-97.
- Beer, Stafford. (1966). Decision and Control. New York: John Wiley and Sons.
- Beer, Stafford. (1979). The Heart of Enterprise. New York: John Wiley and Sons.
- Gharajedaghi Jamshid. (1985). Toward a Systems Theory of Organization. Seaside, CA: Intersystems Publications.
- Hostetter, Gene H., Clement J. Savant, and Raymond T. Stefani. (1982). Design of Feedback of Control Systems. New York: Holt, Rinehart and Winston.
- Klir, George J. (1991). Facets of Systems Science. New York: Plenum Press.
- Leonard, Allena. (no date). "The Viable System Model in Aid of Quality and Productivity." Viable Systems International, Oakland, CA.
- Pearson, John M., Larry B. Rainey and Robert S. Tripp. (1985). "The Application of Cybernetics to the Air Force Logistics Command (AFLC) Command, Control, Communication and Intelligence System." CYBERNETICA, Vol. 28, No. 2: 145-150.
- Rainey, Larry B. (1991). "The Application of Cybernetics to the Ground Segment for Space Operations." Ph.D. dissertation, the Union Institute.
- Schmidt, Stephen R. and Frank D. Watson. (1990). Quality Through Leadership: TQM In Action. no publisher.
- Schoderbek, Peter P., Asterios G. Kefalas and Charles G. Schoderbek. (1975). Management Systems: Conceptual Considerations. Dallas: Business Publications, Inc.

AD-P006 923



Total Quality Management: A Unique Challenge
For Behavioral Science Researchers

J.L. Mitchell, Ph.D.
J.R. Knight, M.A.
McDonnell Douglas Corporation
San Antonio, Texas

92-17656



Abstract

This essay examines the unique new opportunities for behavioral scientists to develop innovative study methods which can be used to systematically collect data to create and support the implementation of Total Quality Management (TQM) programs in industry and government. The essence of TQM involves respect for personal choice and the participation of every level of workers in decisions which impact the organization. Typically, structured questionnaires developed by behavioral scientists are designed to be easy to summarize and generally lack any mechanism for personal comments, ideas, and suggestions. New open-ended, interactive, and iterative survey methodologies are needed to enhance creativity, encourage feedback, and foster a sense of participation.

Total Quality Management (TQM) is a Department of Defense (DoD) initiative for improving performance at all organizational levels. In recent years, DoD has extended this program to defense industries as an essential part of their TQM efforts; the program is also viewed as "an operating philosophy driven by customer satisfaction, total quality, continuous improvement, work force involvement, and a willingness to change" (MDMSC, 1991:2). In the aerospace industries, the program is known by various names: Total Quality Management System (TQMS) at McDonnell Douglas; Total Quality Improvement (TQI) at Boeing; Project Paradise at Martin Marietta; and TQM at General Dynamics (*Ibid.*). For the U.S. Navy, it is Total Quality Leadership (TQL), which is defined as a management system to improve productivity (Shettel-Neuber, Sheposh, Rosenthal, and Heller, 1991).

What is Quality and TQM?

Historically, the emphasis on worker involvement and first-time quality in the workplace has emerged from a number of different trends in research and business. In the early 1970s, the focus was on "job enrichment" with the aim that more interesting jobs should lead to higher productivity. Dr. Michael Beer, manager of organizational research and development for Corning Glass noted, "In the past, we have vastly underestimated what a person was capable of. Now we give them an opportunity to grow into the job" (*U.S. News and World Report*, July 17, 1972:54). By the mid-1970s, there was more emphasis on "job satisfaction" and performance feedback for workers as prerequisites for productivity. In discussing work with a coal-mining company and the United Mine Workers, Professor Gerald Susman of Pennsylvania State University observed, "The aim is not so much to redesign the jobs themselves as to permit miners to make as many decisions as they can on

their own, as well as to train the supervisors to make decisions in a more participatory way--to ask advice of their subordinates on how to do things" (*U.S. News and World Report*, Sept. 27, 1976:88). Another aspect of this trend was to link incentives with improvements in productivity, as with the Scanlon Plan (sometimes referred to as "gain-sharing"). Such programs often led to higher production and decreased absenteeism among production line workers (*U.S. News and World Report*, Dec. 26, 1977:85). Unfortunately, another trend was the changing ratio of production workers to nonproduction workers; in 1947 there were 5 production workers to every nonproduction worker in industry where by 1983 this ratio was only 2 to 1 (*U.S. News and World Report*, May 2, 1983:70). Since that time, there has been a major effort to eliminate white collar jobs thus reducing the number of managers and supervisors involved in direct worker contact. Such reductions necessitate a major change in management philosophy and style.

Armand Feigenbaum, recognized by many as the father of TQM, proposed that total quality was basically a new way of managing; he maintained that successful companies focus on customer satisfaction and bring to bear all the resources of the organization to achieve and maintain product quality (Feigenbaum, 1983). Deming, another pioneer in the field, emphasized the need to examine organizational processes with the view of maintaining and enhancing quality; he advocated the use of statistical process controls, planning for quality, extensive training of personnel, and removal of impediments to workers' pride in their work (Deming, 1982). Juran (1988) recommended that companies establish task teams to solve problems and identify opportunities to improve quality; quality councils were created to plan for yearly improvements in quality as a normal organizational activity. All these proposals have come together in the current explosion of active TQM programs.

Problems in Implementing TQM Programs

Any new operating philosophy or innovative management system is likely to be difficult to implement due to the inherent conservatism and resistance to change found in most bureaucratic organizations. Indeed, there have been several recent reports of problems which organizations have encountered when attempting to implement TQM.

Clark (1990) describes an effort to initiate a TQM program in an Air Force Research and Development (R&D) organization through use of a new Methodology for Generating Efficiency and Effectiveness Measures (MGEEM; see Tuttle and Weaver, 1986). Although the program had top management support, the reaction of the R&D scientists was generally negative. He notes that 20 months after the program began, a survey (94 of 380 persons responding) indicated that 80 percent of respondents felt that TQM/MGEEM was of "No Value" or "Some Value" with only 20 percent indicating more positive responses (Clark, 1990:461). Clark concluded that the program was rejected because people in the laboratory did not have a sense of ownership in the program and that MGEEM did not receive a fair test. He observed that "trying to introduce TQM without considering the behavioral dynamics of the organization significantly reduces the chances for success" (Ibid.:460).

Shettel-Neuber and co-workers recently studied U.S. Navy efforts to implement TQM. Questionnaires were administered to 290 employees of a Navy activity to identify major

impediments to implementing TQL and to assess the characteristics of individuals who were supportive of the program. These researchers concluded that the "strongest impediments to successful implementation were associated with a climate not consistent with TQL principles" (Shettel-Neuber, et al., 1991). In the organization studied, there was a poor fit between current organizational policies and TQL philosophy. In general, efforts to implement TQL programs in the Navy "have fallen short when they have adopted a short-term, programmatic, and segmented approach" (Ibid.).

The Challenge for Behavioral Scientists

The real challenge for behavioral scientists interested in TQM is the development of new methodologies which can be used to gather data for statistical analysis which will also support and enhance employees' sense of participation and ownership of the program. This is a significant challenge since most behavioral scientists are used to designing survey instruments and questionnaires more for ease in administration and for standardized data collection and have tended to avoid any narrative comments which have to be individually interpreted. Typical methods are designed to facilitate data entry and summarization of responses; write-in comments, if permitted at all, are usually ignored or, at best, are just read and filed away.

The very nature of most of our behavioral science questionnaires and survey instruments are so formal and structured that they are not conducive to any kind of interactive feedback. They tend to be very precisely worded to prevent misunderstanding that they are written at a very advanced reading level; as a result, many are misinterpreted or misunderstood. Our emphasis is generally on quantifying responses (ordinal, or at best, interval scales) to facilitate analysis; we seldom if ever are concerned with what impact our questionnaires have on the attitudes of the individual respondent.

In a TQM environment, this has to change. We have to devise new methodologies which will permit incumbents to feel that their attitudes, ideas, and beliefs are important to the organization, and to facilitate their providing feedback on management initiatives and programs. This challenge may be facilitated somewhat by the recent development of computer-based survey administration techniques where the survey instrument is on a floppy disk which can be distributed quickly and easily world-wide for administration on standard Air Force PC hardware (Hudspeth, 1991). Such survey administration software is just a first step; it opens the way for more sophisticated interactive survey software which could be to some degree personalized for each job incumbent. Interactive software could be developed which would react to the individual's responses in such a way as to make the incumbent feel that his or her responses were important to the Air Force and to functional managers.

Computer systems, modem networks, electronic message systems, bulletin boards, and other technological innovations could all be used to build an interactive communication system which could be used for TQM data collection as well as for providing feedback and building a sense of participation and TQM ownership on the part of all members of the organization. The potential for using such hardware and software systems in this way is clearly not a communications or software innovation in that such uses are exactly what such

systems were designed to do. They would, however, be significant management innovations; in most organizations such software has not yet been used as a way to implement TQM.

It is a legitimate role for behavioral scientists to devise the survey instruments, the implementation plans, and the TQM standards through which these kinds of user-friendly, organizationally-supportive, TQM-oriented systems could be brought to bear on such issues. In some instances, such efforts may be extremely complex if we are to optimize the positive impact of such new technology on the attitudes and behaviors of job incumbents. As Hudspeth (1991) points out, there is a considerable amount of Human Factors research which still needs to be done to find the best ways to design an automated survey system. It is equally important that parallel TQM research should be done so that we may assess how incumbent attitudes toward management and the TQM effort can be impacted by the creative ways in which new technologies and software are used within our organizations.

It is imperative that behavioral scientists should use their expertise, their knowledge, and their influence, to assure that such emerging interactive systems are designed in such a way as to preserve the privacy and rights of the individual while encouraging maximum participation and feelings of ownership and belonging to the organization. This will be a difficult and sometimes complex problem for behavioral scientists, but it is an unparalleled opportunity for them to exercise those essential principles of psychological science to instill a greater sense of self-worth for both individuals and the organization as well.

The Essence of TQM

Reduced to its most essential essentials, TQM is basically about how to deal with people so as to enhance their work and their participation in the organization. This does not always have to be a difficult management strategy or complex psychometric measurement problem. This point was driven home to us during the last Psych in the DoD conference (April 1990), when the two of us were sitting in the USAFA snack bar with Colonel Frank Pinch, then Director of Personnel Selection Research & Second Careers, Canadian National Defence Headquarters, Ottawa. Colonel Pinch had just been alerted to his special assignment to a working group of sixty colonels who would develop a plan for an impending (and very substantial) reduction-in-force for CF officers. We were discussing how to go about such planning so as to insure that the TQM principles of involvement and participative management would be preserved.

Colonel Pinch observed that if it were up to him, if he had to do the whole thing himself, he would simply write to every officer outlining the problem (numbers to be reduced), and ask what each wanted to do.

This simple, direct, honest, and straightforward approach to a very difficult problem seems to capture the essence of TQM philosophy. You don't have to equivocate, deliberate, philosophize, or strategize. All you really have to do is to inform people of the problems, of the decisions to be made, answer their questions if you can, and listen to their ideas, suggestions, and decisions. That's really what TQM is all about. As behavioral scientists, that's what we should be making happen in our respective organizations.

References

- Clark, Herbert J. (1990). Starting a TQM program in an R&D organization. Proceedings of the 32nd Annual Conference of the Military Testing Association (pp. 460-465). Orange Beach, AL: Naval Education and Training Program Management Support Activity.
- Deming, W. Edwards (1982). Quality, productivity, and competitive position. Cambridge, MA: Massachusetts Institute of Technology Press.
- Feigenbaum, Armand (1983). Total quality control (3d Edition). Cincinnati, OH: Association for Quality and Participation.
- Hudspeth, D.R. (1991). Research productivity tool for job inventory administration. Proceedings of the 33rd Annual Conference of the Military Testing Association. San Antonio, TX: Armstrong Laboratory Human Resource Directorate and the USAF Occupational Measurement Squadron.
- Juran, J.M. (1988). On planning for quality. Cambridge, MA: Productivity Press.
- Kanji, G.K. (1990). Total quality management: The second industrial revolution. Total Quality Management, 1(1):3-12.
- McDonnell Douglas Missile Systems Company (1991). TOMS awareness training: Reference guide. St. Louis, MO: McDonnell Douglas Missile Systems Company.
- Shettel-Neuber, M.J., Sheposh, J.P., Rosenthal, M., and Heller, C. (1991). Impediments to TQL implementation within a U.S. Navy activity. Proceedings of the 33rd Annual Conference of the Military Testing Association. San Antonio, TX: Armstrong Laboratory Human Resource Directorate and the USAF Occupational Measurement Squadron.
- Staff. (1972, July 17). The drive to make dull jobs interesting. U.S. News and World Report, pp. 50-54.
- Staff. (1976, September 27). Why millions hate their jobs--And what's afoot to help. U.S. News and World Report, pp. 86-88.
- Staff. (1977, December 27). How to keep workers happy on the job. U.S. News and World Report, pp. 85-86.
- Staff. (1983, May 2). As business copes with "White-Collar Bloat". U.S. News and World Report, pp. 69-70.
- Tuttle, T.C., Weaver, C.N. (1986, November). Methodology for generating efficiency and effectiveness measures (MGEEM): A guide for Air Force measurement facilitators (AFHRL TP-86-36, AD A174 574). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.



Student Attitudes Toward Goal Interdependence Characterizations of Competition and Cooperation

R. J. Grisham, Ph.D.
United States Air Force Academy

92-17657



This study examined the interrelationship of student attitudes toward goal interdependence (GI) characterizations of competitive and cooperative educational settings. Consistent with goal interdependence theory, a low positive correlation was found between attitudes of competition and cooperation. Student need for social comparison was positively correlated with student attitudes of competition and cooperation. Student helpfulness attitudes were unrelated to attitudes of competition and strongly and positively correlated with cooperation. Implications for education are discussed.

There are hundreds of studies comparing the relative effects of competitive and cooperative goal structures in classroom settings (Johnson & Johnson, 1989; Slavin, 1980). Many of these theories suggest that competition and cooperation are inversely related along a prosocial continuum with competition representing the antisocial and cooperation the prosocial ends (Deutsch, 1949; Pepitone, 1980). Other research has suggested that competition and cooperation are independent concepts (Johnson & Norem-Hebeisen, 1979). These mixed findings seem to result from a single problem: classroom goal structures and attitudinal scales often confound competition and cooperation with highly correlated yet conceptually distinct concepts such as rivalry and helping. Recent research efforts have begun to test for the individual effects of small, well-defined aspects of goal structures in order to correct for this problem (Sharan, 1980; Slavin, 1980). Attitudinal scales, however, have not limited their measures to narrow, well-defined concepts. The mingling of characteristics has likely obscured and/or distorted the relationships of competition and cooperation, to each other and to a wide spectrum of behaviors and other outcomes. The major aim of this study is to examine the interrelationship of student attitudes toward competition and cooperation using strict goal interdependence characterizations. A second and equally important purpose is to examine the relationships between competition and cooperation with helpfulness and need for social comparison.

Goal interdependence refers to situational conditions, behaviors, or dispositions that create perceptions of goal interconnectedness between or among individuals. The mere presence or absence of goal interdependence leads to perceptions of how much control one wields. Competition and cooperation refer to a perceived presence of goal interdependence; and consequently, to a perception of shared or joint control over one's own outcomes as well as the outcomes of others with whom one is interconnected. In goal interdependence situations, people perceive themselves as exercising only partial control over their own outcomes and partial control over the outcomes of others as well. As these characterizations suggest, it is not the existence or non-existence of mutual control or reliance that distinguishes these concepts; instead, it is the nature of their interdependence. However, the presence of goal interdependence does distinguish competition and cooperation from individualism, and suggests that attitudes toward competition and cooperation may be positively correlated.

Goal interdependence theories suggest that competition exists when goals are negatively linked in such a way that the probability of one person attaining his or her goal is negatively correlated with the probability of others attaining their goals (Deutsch, 1982). In contrast, cooperation exists when goals are positively linked in such a way that the probability of one person attaining his or her goal is positively correlated with the probability of others attaining their goals. As these characterizations suggest, each of these concepts can be viewed as a correlation between the probability of one person attaining his or her goal and the probability of others attaining their goals. Competition represents a negative correlation and cooperation, a positive correlation.

Helping, in contrast, is characterized by goal dependence in which one actor is dependent upon another in order to attain his or her goal. Helping initiates a cycle of dependency involving debt and repayment; and due to its comparative indeterminance the relationship continues across time to new situations (Gouldner, 1960). Research suggests that cooperation and helping should be strongly

correlated and that competition and helping should be inversely related (Deutsch, 1949; Johnson & Johnson, 1989). However, helping and competition may not be antithetical concepts. In fact, most competitive situations are characterized by cooperative competition involving cooperation within one's group and competition with outgroups (Sherif, Harvey, White, Hood, & Sherif, 1988; Slavin, 1980). In these situations, participants often help one another prepare for competition. A preference for competition in the strict sense provides little information about preferences for helping unless competition is confounded with rivalry, thus creating a continuum of prosocial behavior.

The relationship between competition and the need for social comparison has received considerable theoretical and empirical support. Several studies have found that as the availability of socially comparative information increases so does competitive behavior (Dakin & Arrowood, 1981; McClintock & McNeel, 1967). In contrast, the relationship between cooperation and social comparison has not received much theoretical or empirical attention. Cooperative situations are usually structured to minimize comparative information, a manipulation that may operate to avoid competition rather than to structure cooperation. Most theories of cooperation involve principles of equity, equality, and other rules of fairness (Deutsch, 1982) all of which presuppose the availability of socially comparative information. For example, people must have information available regarding the outcomes and efforts of both self and others in order to determine whether their outcomes are equitable or not. The same information is required to make determinations regarding equality. However, unlike competitive tasks which by their very nature require comparative information in order for people to perform, cooperative tasks do not require comparative information and can be performed without contact with other people. This analysis suggests that competitive and cooperative outcomes (products) require socially comparative information, whereas only competitive tasks require this information.

Method

A set of five questionnaires were given to 150 students enrolled in Introductory Psychology at the University of Colorado. The average age for subjects (69 female and 81 male) was 19.01 years with a range from 18 to 26 years. The majority of subjects were enrolled or intended to enroll in either the School of Arts and Science (59%) or the Business School (30%) and were mostly freshmen (49%) and sophomores (37%). The study lasted approximately 90 minutes.

Goal Interdependence Scales. Competition and cooperation scales were constructed for this study by adhering strictly to the conceptualization of goal interdependence theory (Deutsch, 1949, 1982; Johnson & Johnson, 1989; Pepitone, 1980). This instrument was composed of 78 items. The competition scale consisted of 24 items, the cooperation scale had 26 items, and an individualism scale contained 28 items. Each global scale (that is competition, cooperation, and individualism) had two major subscales relating to grades and tasks. Each of the major subscales (grade and task) were divided into minor subscales that measured student attitudes toward different aspects of the various interdependence conditions. These minor subscales, each of which had from two to seven items, were: 1) global attitudes, for example, "I like to work on school projects that require both my own efforts and the efforts of other students in order to complete them;" 2) motivation, "I feel motivated when my grades are based on my own efforts and no one else can influence them;" 3) relationships, "I like the relationships that develop between students when they try to outperform each other on a task;" and 4) instrumentality, "I learn my school work quickly when I'm trying to outperform others." Students responded to each item by indicating the extent to which they agreed or disagreed with each statement on a six-point, forced-choice, Likert-type scale. The Cronbach alphas for the global scales were: competition scale, .95; cooperation scale, .91; individualism scale, .92.

Helpfulness Scale. The helpfulness scale asked students about their attitudes toward helping behaviors in educational settings. The scale was composed of sixteen items referring to helping tasks and represented four content areas or subscales: 1) global preferences, 2) motivation, 3) relationships, and 4) instrumentality. The Cronbach alpha for this scale was .91.

Need for Social Comparison. This scale was composed of six items that referred to a need for students to compare their grades and performances with the grades and performances of other students. For example, "I feel a strong need to know how others have performed on a test." The Cronbach alpha for this scale was .87.

Results

Goal Interdependence Scales. Two approaches were taken to develop goal interdependence scales and to determine the relationships between them. All items for the 19 subscales were factor and item analyzed to develop a statistically and conceptually relevant group of subscales. Next, these subscales were factor analyzed. Several extraction and rotation procedures were employed all of which gave essentially the same solution. The second approach was to develop a priori or additive scales by adding together the scores of all competitive items, and to do the same for the cooperation and individualism scales. Both the factor analytic and a priori analyses gave almost identical results for the global scales. The results from the a priori scales will be reported in order to discuss the grade and task subscales.

The competition and cooperation scales had a low positive, yet significant, correlation ($r = .19, p < .05$), the competition and individualism scales were uncorrelated, and the cooperation and individualism scales were significantly and negatively correlated ($r = -.47, p < .01$).

When the a priori global scales were decomposed into grade and task subscales, a different pattern of relationships was found (see Table 1). Grade and task subscales, within each global scale, had high positive correlations. Competitive grades had a significant positive correlation with cooperative grades and a significant negative correlation with individualistic grades. In addition, cooperative grades had a significant negative correlation with individualistic grades. This pattern of results appears to contrast competition and cooperation with individualism along a goal interdependence continuum. That is, competitive and cooperative grades represent conditions of goal interdependence whereas individualistic grades represent a condition of no goal interdependence. However, this pattern was not replicated for interdependent tasks. Competitive and cooperative tasks had a low positive and insignificant correlation; competitive and individualistic tasks were uncorrelated; and cooperative and individualistic tasks had a significant negative correlation.

Table 1. Intercorrelations of Goal Interdependence Global Scales, and Grade and Task Subscales (* $p < .05$; ** $p < .01$).

	Comp		Coop		Ind
	grade	task	grade	task	grade
Competition					
grade					
task	.85**				
Cooperation					
grade	.27**	.25**			
task	.10	.16	.71**		
Individualism					
grade	-.25**	-.08	-.39**	-.24**	
task	.01	.03	-.37**	-.54**	.57**

Need for Social Comparison. The GI global scales displayed a different pattern of relationships with need for social comparison than did their decomposed grade and task subscales (see Table 2). The results for the global GI scales supported the hypotheses that competition and cooperation are positively correlated with the need for social comparison and that competition correlates more highly than cooperation with social comparison, $t(141) = 2.12, p < .05$. However, the grade and task subscales showed that competitive grades ($r = .39, p < .01$) and tasks ($r = .34, p < .01$) were both positively correlated with social comparison, but only cooperative ($r = .23, p < .01$) and individualistic ($r = -.18, p < .05$) grades, not tasks, were correlated with social comparison. This pattern of results suggests that the need for social comparison is activated whenever the outcomes of an activity are considered, regardless of the nature of the interdependency. However, this conclusion needs further investigation because the low, yet significant, negative correlation between individualistic grades and social comparison may be due to the large sample size. In contrast, only

competitive tasks, which by their very nature require comparative information in order to perform,

Table 2. Correlations of Need for Social Comparison (SC) and Helpfulness (H) with Goal Interdependence Global Scales and Grade and Task Subscales (* $p < .05$; ** $p < .01$).

	SC	H
Competition	.36**	.00
grade	.39**	.00
task	.34**	.00
Cooperation	.20*	.58**
grade	.23**	.39**
task	.14	.64**

seem to activate the need for social comparison.

Helpfulness. The results for the global goal interdependence scales supported the hypotheses that competition is uncorrelated with helpfulness ($r = .00$) and that cooperation is positively and significantly correlated with helpfulness ($r = .58, p < .01$). The GI grade and task subscales displayed a similar pattern of correlations with helpfulness. As expected, the competitive grade and task subscales were uncorrelated with helpfulness (r 's = .00), and the cooperative grade and task subscales were positively and significantly correlated with helpfulness ($r = .39, r = .64$, respectively; p 's < .01).

Discussion

The results suggest that student attitudes about competition and cooperation are essentially independent concepts. Independence, as applied here, means that student ratings of competition provide little information about how students rate cooperation; and therefore, that competition and cooperation do not represent opposite ends of a continuum. The results definitely do not support the hypothesis that competition and cooperation are inversely related as has been suggested in previous research (Deutsch, 1982; Pepitone, 1980). Deutsch and Pepitone have suggested that a prosocial continuum separates people's perceptions of competitive and cooperative situations. When the effects of the prosocial continuum are removed, as was done in the GI attitudinal scales, the relationship between competition and cooperation is not negative. In fact, the data consistently show, across factor analytic scales and a priori scales that the relationship is a low positive one. The shared variance was consistently around 5%.

Because the items measuring competition and cooperation adhere strictly to goal interdependence theories, it is likely that the positive relationship results from their characterizations as goal interdependence conditions. Although both conditions require people to rely on one other for their outcomes or to complete a task, the nature of their interdependence, positive versus negative, is dissimilar and therefore likely to keep the correlation low. It is likely, although it was not specifically tested, that the relationship would become negative as the characterizations include items representing bipolar contrasts such as the rivalry-helpfulness prosocial continuum.

One of the most interesting findings is that student attitudes toward competition and helpfulness are unrelated. Much of the previous research has compared the amount of helping found in cooperative and competitive situations (Johnson & Johnson, 1989; Weigel, Wiser & Cook, 1975) and has consistently shown that people are more helpful in cooperative situations. This strategy has failed to examine competitive conditions that might increase helping behaviors like those suggested in studies of cooperative competition (Sherif, Harvey, White, Hood & Sherif, 1988). As Mead (1937) suggests, the intention of competitive activities is not necessarily to obstruct the efforts of others to achieve, to reduce helping and sharing, or to increase antisocial tendencies as found in other studies (Miller & Hamblin, 1963). Situations can be structured so that the primary goal of competitive activities is to encourage individuals to help one another. Competition can be structured so that ingroup members help one another prepare for outgroup competition as is found in the Teams-Games-

Tournament cooperative learning method (Slavin, 1978), or so that the competitive activity itself is perceived to help people sharpen their skills for a future activity as is found in intra-squad scrimmages for a sports team.

The social comparison data support an earlier finding (Liebrand & McClintock, 1988) and theories suggesting that cooperative outcomes (grades) elicit a need to compare one's own outcomes with the outcomes of others (Deutsch, 1982; Kagan, 1977). It is believed that cooperative comparisons help people to decide if an interaction has been fair using various rules such as equity, equality, group enhancement, or need. This finding also suggests that previous studies that found competitive behaviors to increase as the amount of socially comparative information available to subjects increases (McClintock & Messick, 1966; McClintock & McNeel, 1967) should expand future studies to include the systematic testing of different purposes for comparison. The results also suggest that social comparison is an integral part of competitive activities, which by their very nature require comparative information to perform the activity, but is not integral to cooperative or individualistic tasks.

References

- Dakin, S., & Arrowood, A. J. (1981). The social comparison of ability. Human Relations, 34, 89-109.
- Deutsch, M. (1949). A theory of cooperation and competition. Human Relations, 2, 129-152.
- Deutsch, M. (1982). Interdependence and psychological orientation. In V. J. Derlega & J. Grzelak (Eds.), Cooperation and helping behavior: Theories and research (pp. 15-42). New York: Academic Press.
- Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. American Sociological Review, 25, 161-178.
- Johnson, D. W., & Norem-Hebeisen, A. A. (1979). A measure of cooperative, competitive, and individualistic attitudes. Journal of Social Psychology, 109, 253-261.
- Johnson, D. W., & Johnson, R. T. (1989). Cooperation and competition: Theory and research. Edina, Mn: Interaction Book Company.
- Kagan, S. (1977). Social motives and behaviors of Mexican-American and Anglo-American children. In J. L. Martinez (Ed.), Chicano psychology (pp. 45-86). New York: Academic Press.
- Liebrand, W. B. G., & McClintock, C. G. (1988). The ring measure of social values: A computerized procedure for assessing individual differences in information processing and social value orientation. European Journal of Personality, 2, 217-230.
- McClintock, C. G., & McNeel, S. P. (1967). Prior dyadic experiences and monetary reward as determinants of cooperative behavior. Journal of Personality and Social Psychology, 5, 282-294.
- McClintock, C. G., & Messick, S. P. (1966). Reward feedback as a determinant of competitive and cooperative game behavior. Journal of Personality and Social Psychology, 4, 606-643.
- Mead, M. (1937). Cooperation and competition among primitive peoples. New York: McGraw-Hill.
- Miller, L. K., & Hamblin, R. L. (1963). Interdependence, differential rewarding and productivity. American Sociological Review, 28, 768-778.
- Pepitone, E. A. (1980). Theoretical orientation. In E. A. Pepitone (Ed.), Children in cooperation and competition: Toward a developmental social psychology (pp. 67-104). Lexington, MA: Lexington Books.
- Sharan, S. (1980). Cooperative learning in small groups: Recent methods and effects on achievement, attitudes and ethnic relations. Review of Educational Research, 2, 241-271.
- Sherif, M., Harvey, O. J., White, B. J., Hood, W. E., & Sherif, C. W. (1988). The Robbers Cave experiment: Intergroup conflict and cooperation. Middletown, CT: Wesleyan University Press.
- Slavin, R. E. (1978). Student teams and academic divisions. Journal of Research and Development in Education, 12, 39-49.
- Slavin, R. E. (1980). Cooperative learning. Review of Educational Research, 50, 315-342.
- Weigel, R. H., Wiser, P. L., & Cook, S. W. (1975). The impact of cooperative learning experiences on cross-ethnic relations and attitudes. Journal of Social Issues, 31, 219-243.



Gender and Attitudes as Predictors of Video Game
Performance and Satisfaction

Brittany J. Thurber
David B. Porter, D.Phil.
Anne C. Parker
US Air Force Academy

Gender and attitudes toward cooperative and competitive environments were examined as predictors of video game performance. Subjects participated in groups of four, playing as pairs that were rotated each of the five trials. They were motivated toward an interdependent goal, a prize based on group score. Although gender and attitude did not predict performance directly, their interaction explained 28 percent of performance variance. Significant differences between male and female scores were found only in the competitive groups where males performed better than females. A post-task survey revealed that males rated their own performance, as well as that of others more highly than did females. In this study the relationship between attitude and performance appeared to be moderated by gender.

92-17658



In his theory of individual tension systems, Morton Deutsch (1949, 1962) conceptualized two distinct goal structures. He defined cooperative situations as those in which each individual's goals are linked, creating positive relationships among all members of the group. An individual can only attain his or her goals if other group members attain their goals as well. Thus, it is beneficial for group members to aid one another and manage conflicts constructively. By working in cooperative situations, people realize they are successful as a team (Tjosvold, 1988), and as such seek results that benefit all involved.

In contrast, a competitive situation is one in which the relationships between individual goal attainments are negative. Individuals find themselves in win-lose situations in which their personal success is incompatible with the success of other group members (Tjosvold, 1988). It is not too surprising that in such situations, the following behaviors are often observed: coercion, threat, refusal to facilitate the achievement of another's goals, or outright attempts to interfere with the performance of others (Johnson, Maruyama, Johnson, Nelson & Skon, 1981).

Individuals differ in their preferences for these alternative reward structures. Theoretically these differences fall along a single dimension, with individuals falling along a single continuum. Reward structures influence group processes, intragroup communication, and the effectiveness of training and educational programs. Individual performance as well as satisfaction can be affected profoundly by these alternative types of reward structures. Previous studies have found that emphasis on interpersonal competition often significantly degrades the quality of education in North American classrooms (Johnson & Johnson, 1975). Some educators' dissatisfaction with academic achievement has led them to look for alternative reward systems which promote higher achievement. Tjosvold (1988) found that introducing cooperative goals to academic groups increased the exchange of information, sharing of resources, productivity, and morale. Porter, Bird, and Wunder (1991) found that individual attitudes interacted with reward structures to predict much of the variance in group performance. Cox, Lobel

and McLeod (1991). found that ethnic diversity tended to decrease the competitive orientation of the group.

The present study examines gender and attitudes toward cooperation and competition as predictors of performance on a video game. Groups of four (a size found to accentuate cooperative learning effects (Hagman & Hayes, 1986), with distinctive attitudes worked together in a cooperative environment. It was expected that the use of an interdependent goal would encourage teammates to share information and strategies. Our hypothesis was that groups with preferences for cooperative environments would perform better in this study. Wrightsman, O'Connor, and Baker (1972), had found that female pairs tended to be less cooperative than male pairs; we wondered if this would hold true for Air Academy freshmen who had undergone extensive training and experience in group oriented activities.

Method

A survey was given to 230 freshmen cadets enrolled in General Psychology at the U.S. Air Force Academy. The survey assessed subjects' preferences for cooperative or competitive environments as well as soliciting certain demographic information. From these data, 48 subjects with little computer game experience were selected. The subjects (24 female and 24 male) were between the ages of 18 and 21 and volunteered to participate. The experiment lasted about 35 minutes and consisted of playing a short sequence of the Nintendo game five times and then answering five questions.

Gauntlet is a two-player video game in which each player controls a separate character. The goal is to fight through a series of mazes gathering gold and other objects along the way. To proceed to the next maze, both subjects' characters must pass through fixed exit portals. Characters could sustain only a certain level of injury from the video villains lurking in the maze before dying, after which their character disappeared from the screen and received a score of zero for that trial. During the experiment, subjects were halted after they completed three mazes and their score for the trial was computed. Each individual's score was equal to their remaining life score plus ten times the number of gold pieces collected. It was important for both players to collect gold, avoid or destroy enemies, and finish quickly.

The 48 subjects were divided into four-person groups with two females and two males in each group. There were three types of groups: those with cooperative attitudes, those with competitive attitudes, and those with individuals with different orientations. These "mixed" groups had one member of each gender that had a cooperative preference and one with a competitive preference. Each four-person group performed the experiment separately. Each individual played five trials of the game, paired with another subject each time. Every trial, the subjects either changed partners or game characters; the fifth trial was a repeat of the first. The order of pairings was counter-balanced across experimental groups. The reward structure for all groups was interdependent (i.e., the most successful group of four individuals was to be our guest at a pizza party). Subjects were encouraged to help each other, within pairs and between pairs, throughout the experiment. Several dependent measures were taken. Individual scores were calculated and then pair and group scores were determined by adding individual scores. The post-experiment survey provided a measure of satisfaction, a self-rating of performance, and a rating of others' performance, on ten-point Likert scales.

Results

The data were analyzed at three levels: individual, pair, and group. The individual data (performance scores and survey results) were analyzed to identify predictors of individual performance and satisfaction. The pair data provided additional information on interactive gender effects. The analysis of the group data was intended to show how attitudes within the groups affected overall group performance.

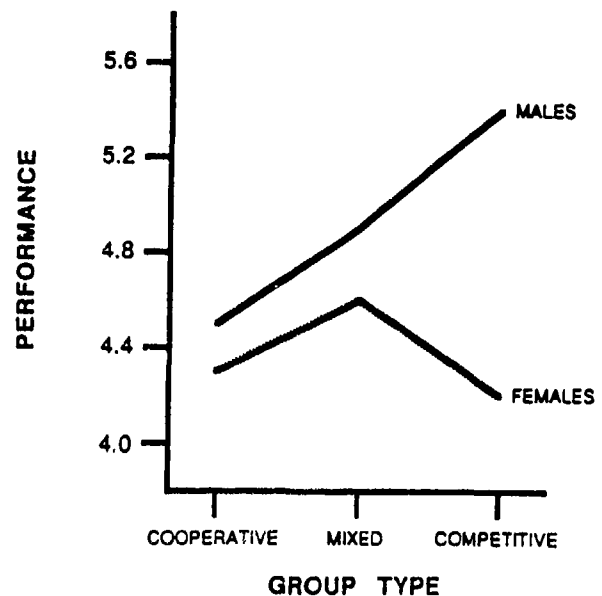
The subjects' raw performance data were coded into nine distinct categories. The distribution of these scores was normal with a mean of 4.65. Survey responses were also normally distributed. The mean rating of others' performance (10 point Likert scale) was 5.25, and subjects' mean rating of their own performance was 6.17. The mean rating for satisfaction with the experiment was 8.52.

Multiple regression analysis was used to discern the effects of gender and attitude on performance. Separately, gender and attitude contributed very little to the variance in performance, but their interaction explained 27.7% of the variance ($F(3,44)=5.63$ $p<.002$). Performance and the rating of their own performance explained 35.4% of the variance in subjects' rating of others ($F(2,45)=12.33$ $p<.001$). Of the variance in subjects' rating of their own performance, 31.3% was explained by performance, gender, and the rating of others' performance ($F(3,44)=6.70$ $p<.001$). Good performance positively affected subjects' rating of others but had a negative effect on self-rating. Those who did best on the task had a tendency to rate others more positively and themselves somewhat less positively. Gender affected both ratings; males tended to rate both themselves and others more positively, while females rated themselves negatively and others even worse. Gender and performance were found to predict individual performance ($F(3,44)=5.63$, $p<.01$). The regression equation was found to explain 28 % of the variance in performance:

$$\text{Performance} = -.31 \text{ Attitude} + .12 \text{ Gender} + .68 \text{ Interaction}$$

The interaction was the only term in the equation which was significant ($t(1,44)=2.52$, $p<.05$). There were no differences in either the cooperative or mixed groups. However, male and female performance within the three groups comprised of individuals with competitive attitudes clearly differed. The six males in the three competitive groups scored significantly higher than the six females ($x_m=5.39$ and $x_f=4.19$).

Data for each of the pairs across all five trials were analyzed to identify which variables most affected the pairs' scores. Mixed gender pairs performed slightly better than female pairs. In the early trials the mixed pairs were also slightly better than the male pairs, but with practice, males performed better than the mixed pairs. These relationships were only marginally significant. There were also noticeable but decreasing practice effects across trials ($F(4,115)=35.3$ $p<.001$).



Gender performance across groups.

Discussion

Neither gender nor attitude alone predicted individual performance. However, together their interaction predicted a large part of the observed variance in individual performance. Males with competitive preferences performed best, even in an environment intended to encourage cooperation. This result appears to run counter to the findings of Porter, Bird, and Wunder (1991) in which individuals with competitive attitudes performed poorly in cooperative settings. This may have been due to differences between the tasks and the groups in these two experiments. The present task was not nearly as interdependent as the one used in the previous study. Unlike the interaction required by sharing control of a single character required in the Porter, et al study, Gauntlet provided opportunities for individuals to control separate characters and receive individual feedback. Although individual scores were summed for pairs and groups before being recorded, individuals observed their own scores and thus received direct individual performance feedback from the game itself. In the study by Porter, et. al. (1991) pairs of subjects had to interact with one another to control the movements of a single character, thus could not receive separate individual feedback on their performance. The less interdependent task may have allowed subjects with competitive preferences to see the task as being competitive rather than cooperative, hence their higher scores. However, females with competitive attitudes seem to have reacted in just the opposite manner; their level of performance was considerably below that of their less competitive counterparts. Another difference between these studies, was the equal numbers of male and female subjects (the previous study had only one female in each six person group).

Analysis of the survey also provided interesting results. It indicated that males tended to rate their performance as well as the performance of others more highly than did females. The correlations between rating of own performance and the rating of others is notable due to their counter-intuitiveness. Those who performed well tended to rate themselves poorly but others more highly. Apparently, they attributed their high scores to the others in their group, not themselves. Those who performed poorly rated themselves highly and the others poorly; apparently attributing their low score to the rest of the group. This suggests subjects did see the task as being very interdependent but also reiterates the difficulties inherent in the use of self-reports. The relationship found across pairs indicating female pairs performed worse than others, as well as females generally depressed ratings of self and others is consistent with the earlier work by Wrightsman, et al (1972).

The disparity in performance between males and females in the groups with competitive attitudes may be related to the specific combination of attitude and gender that existed during the experiment. The findings of Jago and Vroom (1983) indicated female leaders are only rated positively when they display traditional behaviors. Females with competitive attitudes in this group may not have been displaying the "harmonizing" behavior expected of females in group settings. By not conforming to their expected roles, the females may have alienated their male partners who, in turn, found subtle ways to suppress female performance. Such effects may be largely an artifact of the experimental situation. In fact, a comprehensive meta-analysis by Eagly and Johnson (1990) suggests gender stereotypes often affect laboratory studies using ad hoc groups (such as this one) but in studies of performance in actual organizational settings or with intact groups, gender differences are seldom found.

At a more general level, these results suggest that the relationship between attitude and performance under a particular reward structure can be strongly moderated by gender. In this experiment an attempt was made to establish a cooperative reward structure. One way to make sense of the pattern of results would be to suppose that females perceived the experimental situation as being cooperative while males, focusing on the individual feedback available, interpreted the situation as being competitive (cf. Cox, et al, 1991). In a competitive environment, we would expect performance to be positively related to subjects' preference for competition (which was true of males). In cooperative environments we would expect performance to be negatively related to subjects' preference for competition (which was true of females). Perhaps we have once again found that males and females perceive situations differently and what is perceived as real may become real in its consequences.

References

- Deutsch, M. (1949). A theory of cooperation and competition. Human Relations, 2(2), 129-152.
- Deutsch, M. (1962). Cooperation and trust: some theoretical notes. In M. Jones (Ed.), Nebraska Symposium on Motivation, 10.
- Cox, T.H.; Lobel, S.A. & McLeod, P.L. (1991). Effects of ethnic group cultural differences on cooperative and competitive behavior on a group task. Academy of Management Journal, 34(4). 827-847.
- Eagly, A.H. & Johnson, B.T. (1990). Gender and leadership style: a meta-analysis. Psychological Bulletin, 108(2). 233-256.
- Hagman, J. & Hayes, J. (1986). Cooperative learning: Effects of task, reward, and group size on individual achievement. US Army Research Institute for the Behavioral and Social Sciences Report 10.
- Jago, A.G. & Vroom, V.H. (1983). Sex differences in the incidence and evaluation of participative leader behavior. Journal of Applied Psychology, 67, 776-783.
- Johnson, D.W. & Johnson, R. (1975). Learning together and alone: Cooperation, competition, and individualization. Englewood Cliffs, NJ: Prentice Hall.
- Johnson, D.W., Maruyama, G., Johnson, R.T., Nelson, D., & Skon, S. (1981). Effects of cooperative, competitive, and individualistic goal structures on achievement: a meta-analysis. Psychological Bulletin, 89, 47-62.
- Porter, D., Bird, M., & Wunder, A. (1991). Competition, cooperation and the performance of complex tasks among Air Force cadets. Current Psychological Research and Reviews, 9(4). 347-354.
- Tjosvold, D. (1988). Cooperative and competitive dynamics within and between organizational units. Human Relations, 41, 425-436.
- Wrightsmen, L., O'Connor, J., & Baker, N. (1972). Cooperation and competition: Readings on Mixed-Motive Games.



The 13th Biennial Psychology in the
Department of Defense

Operational Implementation of a Validated Personnel Selection
System for Landing Craft Air Cushion (LCAC) Vehicle Operators

Daniel L. Dolgin, Ph.D.
David R. Street, Ed.D.
Tatree Nontasak, Ph.D.
David J. Blower, Ph.D.
Kathleen Travis, M. A.

Naval Aerospace Medical Research Laboratory
Pensacola, Florida 32508-5700

Abstract

To support the performance-based selection of LCAC operators, we identified and evaluated a battery of personality, cognitive, and psychomotor tests at the Naval Aerospace Medical Research Laboratory. Operational testing of LCAC candidates for selection into training began in May, 1990. Results indicate that several tests are predictive of LCAC training criteria. Fifty LCAC candidates took the test battery as a requirement for entry into the LCAC training program. We collected and analyzed training performance data from those candidates that achieved a passing score on the LCAC test battery. Descriptive statistics and correlation coefficients were derived. Psychomotor test performance demonstrated a significant relationship with the LCAC performance criteria. Results indicate that the use of computer-based testing yields a selection success rate of 96% for LCAC operators completing training. The implications of these findings for LCAC personnel selection are discussed.

Introduction

One of the world's most unique transport vehicles is the Landing Craft Air Cushion (LCAC) vehicle. The LCAC "rides" on a cushion of air generated by large fans that allow it to negotiate both land and water surfaces. The LCAC is similar to a helicopter in that it has six dimensions of motion. Operating the LCAC demands unique perceptual and psychomotor skills. In addition, with a machine as expensive and inherently dangerous as the LCAC, sound judgment and decision-making play an important role.

As training costs escalate and projected plans call for an increased number of LCAC vehicles and crews, selecting candidates who will be successful in the operation of complex machinery becomes more critical. In 1988, the Naval Aerospace Medical Research Laboratory (NAMRL) completed the development of a computer-based performance assessment battery for LCAC. This automated battery assesses basic information processing abilities,

higher order processes, psychomotor skills, time-sharing ability, and personality traits that might predict success in LCAC training. The test battery was based on previous literature and research from pilot selection, e.g., Damos and Gibb (1987); Dolgin and Gibb (1989). A recent report based on an aviator selection model (Blower and Dolgin, 1991) supports the efficacy of computer-based assessment in selection.

A concurrent validity study examining Navy fleet LCAC operator performance on the automated test battery was reported recently (Nontasak, et al. 1990). In those studies, several predictor tests were significantly correlated with measures of success in LCAC training. The reports contain the results of an initial predictive validation of the test battery for LCAC trainees. Descriptive statistics and individual correlations between test measures and LCAC training criteria are presented.

The criterion measure used was pass or fail in LCAC primary training. The underway grade (UG) is the core criterion and represents a composite score reflecting a student's tactical performance in the training hovercraft. We did not relate test battery measures to specific components of LCAC training. Past research in flight training has shown the difficulty of identifying reliable subcriteria embedded within the more global criteria (Dolgin, et al. 1987).

METHODS

Subjects

Prospective LCAC operator candidates ($N = 50$) were selected for training on the basis of current medical examinations and NAMRL computer-based testing results. The operator candidates were informed prior to testing that the results would be used in the final selection decision. The candidates were 20 to 42 years old ($M = 31.19$, $SD = 5.93$). All candidates were male and had successfully completed a minimum of a high school education.

Apparatus

All testing was conducted on Apple IIe microcomputers with control sticks, foot pedals, and throttle. Subjects used a numeric keypad to respond to discrete stimuli. All responses were recorded to millisecond accuracy. For psychomotor tests (PMT), two control sticks (Measurement Systems, Inc., 542) were used for cursor control during the tracking tasks. One control stick was mounted in the center on the forward edge of a standard straight-back metal chair. The other stick (throttle) was mounted on the left edge of the chair. Additional apparatus included rudder pedals patterned after those of a Systems Research Laboratory psychomotor test device. Two Jameco JE 520-AP Voice Synthesizers were used to present the Dichotic Listening (DLT) letter-digit strings over binaural headphones. Further details are included in Blower and Dolgin (1991).

Procedure

All subjects were tested prior to entering LCAC training. Instructions were presented to the subjects on the CRT for each task. Test administrators intervened only to begin the computer program for each task and to answer questions posed by subjects at any time. The test battery took from 2.15 to 2.30 h to administer with a 3-4 min rest period given after each task.

Results

Psychomotor Task and Dichotic Listening Task

Performance on the series of psychomotor multitask conditions correlated with training criteria. Table 1 lists the tests that correlated significantly with training criteria. When performed in combination with the DLT, PMT tasks indicated a high relationship to UG with coefficients ranging from $-.54$ to $-.60$. These correlations were in the expected direction. That is, greater psychomotor tracking error was associated with lower overall UGs.

TABLE 1. Summary of Significant Correlation Values between Test Measures and Training Criteria

Test measure	Underway grades
PMT (error score)	
Stick-DLT	$-.60^*$
Stick-rudder-throttle + LLT	$-.54^*$
One-dimensional Compensatory Tracking	
CDOT	$-.45^*$
CDOT + Digit Cancellation Composite	$.34^*$
Digit Cancellation	$.27$

* $p < .05$, two-tailed

The combined PMT/DLT (stick control) yielded the highest correlation with the UG ($r = -.60$). This relationship was significant at the .001 level. The correlation ($r = -.54$) between the PMT/DLT (stick, rudder, and throttle controls) composite and the UG was also significant at the .001 level.

One-dimensional Compensatory Tracking Digit Cancellation

The average cumulative distance off target (CDOT) error for the final three trials of the task was used because it was most

stable. Average CDOT error was $M = 17.79$, $SD = 5.79$ for those recommended for LCAC training ($n = 36$), the average CDOT for those not recommended was $M = 29.35$, $SD = 10.16$ ($n = 14$). The average CDOT difference between the two groups was significant ($t = -5.07$, $p < .001$). UG correlated significantly with average CDOT ($r = -.45$, $p < .01$). When the one-dimensional compensatory tracking task was performed in combination with the digit cancellation task, the correlation between their composite score and UG was $r = .34$ ($p < .05$). As can be seen by the results, the average CDOT and CDOT/digit cancellation composites correlated significantly with student UG during LCAC training.

The average number of digits cancelled in the Digit Cancellation task explained additional variance in the prediction equation and was used in the final selection system. The average digits cancelled for those recommended for LCAC training was $M = 106.23$, $SD = 20.31$ ($n = 36$). The average digits cancelled for those not recommended for training was $M = 78.59$, $SD = 19.88$ ($n = 14$). The difference between average digits cancelled for the two groups was significant ($t = 4.35$, $p < .001$). The correlation ($r = .27$) between UG and the average digits cancelled was not significant at the .05 level.

Discussion

The psychomotor task performed in combination with the dichotic listening task was significantly related to the pass/fail criterion. These results provide support for the notion that certain abilities are useful in screening individuals for the LCAC training program.

Subjects who perform better (lower error score) on the multitask tests have a greater likelihood of success in the training syllabus. Taken together, these characteristics have proven to be reliable predictors of 1) those who will fail from training, and 2) actual LCAC performance in the primary portion of tactical, underway training. In these times of constrained military budgets, these variables may be crucial in reducing attrition from training.

The current findings support previous research in LCAC personnel selection (Dolgin and Nontasak, 1990) and corroborate a computer-based LCAC selection system. Continued research and evaluation of the valid tests in the battery and follow-up with the successfully screened LCAC-trainee population will provide insight into the long term value of the LCAC selection system in decreasing accidents and improving the quality of LCAC operators in the U.S. Navy.

Acknowledgments

This work was sponsored by NAVSEA (Code PMS 377) through Naval Ocean Systems Center and supported by Naval Coastal Systems Center (Work Unit Number CE02). The views expressed in this article are

those of the authors and do not reflect the official policy or position of the Department of the Navy, Department of Defense, nor the U.S. Government. Volunteer subjects were recruited, evaluated, and employed in accordance with the procedures specified in the Department of Defense Directive 3216.2 and Secretary of the Navy Instruction 3900.39 series. These instructions are based upon voluntary informed consent and meet or exceed the provisions of prevailing national and international guidelines. Trade names of materials and/or products of commercial or nongovernment organizations are cited as needed for precision. These citations do not constitute official endorsement or approval of the use of such commercial materials and/or products. The authors wish to acknowledge the assistance of Peter Collyer for his programming effort and Sylvia Starling for test administration.

References

- Blower, D. and Dolgin, D. (1991). An evaluation of performance-based tests designed to improve Naval Aviation Selection, NAMRL-1363, Naval Aerospace Medical Research Laboratory, Pensacola, FL.
- Damos, D. and Gibb, G. (1987). Development of a computer-based naval aviation selection test battery, NAMRL-1319, Naval Aerospace Medical Research Laboratory, Pensacola, FL.
- Dolgin, D. and Gibb, G. (1989). Personality assessment in aviator selection: Past, Present, and Future. In R. Jensen (Ed.), Aviation Psychology: The International Contribution., Gower Publishing Group, London, 285-319.
- Dolgin, D. Nontasak, T., Gibb, G., & Helm, W. (1987). Instructor pilot evaluations of key naval primary flight training criterion, NAMRL-1331, Naval Aerospace Medical Research Laboratory, Pensacola, FL.
- Dolgin, D. and Nontasak, T. (1990). Initial validation of a personnel selection system for Landing Craft Air Cushion (LCAC) vehicle operators, Proceedings of the 12th symposium of Psychology in the Department of Defense, USAF Academy, Colorado, April, 245-249.
- Nontasak, T., Dolgin, D., and Blower, D. (1991). Performance differences in psychomotor and dichotic listening tests among Landing Craft Air Cushion vehicle operator trainees. Proceedings of the Human Factors Society 35th Annual Meeting, San Francisco, California, 987-990.



The Role of Personnel Testing in

a Changing Air Force

Major Constance L. Riggs

Ms Tina S. Ziebell

Air Force Military Personnel Center

92-17660



Abstract

The Air Force relies heavily on personnel testing as a method of selecting, promoting, and identifying individuals with the necessary aptitudes, skills, and abilities to perform well. The testing function has become increasingly more important as the force draws down—it is imperative to make the best possible personnel decisions in support of mission accomplishment.

The purpose of this paper is to examine the changing role of personnel testing in the Air Force in this new environment. To partially compensate for the diminished numbers of personnel, we are looking at increasing efficiency across-the-board by revising procedures, consolidating functions, and realigning responsibilities. At every opportunity, we are integrating new technologies.

Current Organization

The Personnel Testing Branch is part of the Air Force Military Personnel Center (AFMPC) at Randolph Air Force Base (AFB), Texas, the operational hub for the AF military personnel system. Currently, the Personnel Testing Branch is in the Personnel Measurement Division of the Personnel Operations Directorate. The Branch is responsible for oversight of 600 test control officers (TCOs) who administer 700 different tests to more than 275,000 personnel annually. These tests fall into three categories: Personnel Procurement, Aptitude, and Interest Testing; Personnel Promotion Testing; and Personnel Proficiency Testing. The Branch is also responsible for monitoring research and development of new testing programs, providing guidance on new technology for testing programs, and maintaining accountability and control of test material throughout the Air Force.

The Personnel Testing System provides appropriate instruments to measure aptitudes, knowledge, and other abilities of AF applicants and personnel. Testing saves money and resources by selecting the best people to access, train, and place in specific jobs. It also assists in identifying the most qualified airmen for promotion and identifies members who have reached a level of proficiency required for some special area (for example, a career specialty or a foreign language).

Changing Role of Testing in the Air Force Environment

The dynamic world environment, including decreases in the AF budget and force, has necessitated we do business more efficiently. We have accepted the challenge and are stepping out as discussed below.

The Weighted Airman Promotion System (WAPS) uses promotion tests as one component to compare promotion eligibles with their peers and determine who should be promoted. The WAPS has been in effect for almost 20 years and is well accepted and understood by the Enlisted Force. Because the system has worked well over the years, changes are not easily approved or accepted. Of course, integrity of the system and fairness to individuals are always primary concerns. However, procedural changes to the system have been required to keep pace with events.

During Operations DESERT STORM, FIERY VIGIL, and PROVIDE COMFORT, circumstances dictated we alter our normal approach. When fighting in the Persian Gulf began, testing was suspended for those personnel in the area of responsibility (AOR); however, testing for all others remained intact. As personnel returned from these contingencies, they were given, as an exception to policy, 75 days to study and take care of personal concerns. Under normal circumstances, personnel who are unable to test before or during their TDY are tested as soon as possible upon return. This policy for extended study time created some strange situations handled on a case-by-case basis. After the fighting was over and the peacekeeping force began rotating in and out, the policy was questioned many times. When does the extended study time end? How are these individuals different from others who go to less-than-desirable locations to perform their duties and are given no special considerations?

When Mt Pinatubo erupted in the Philippines, similar circumstances arose. People were evacuated quickly--their lives disrupted. Others were required to stay and help with the evacuation, protect American property, and/or clean up. Again, extra time was given to these personnel and many individual circumstances had to be considered.

As the contingencies settled and the urgency abated, other issues surfaced. Is it fair to continue allowing extra study time to these personnel after the threat of combat has subsided? How is the transition made back to normal testing procedures? Will members be disadvantaged upon return to normal testing policy? As a result, current policy for testing members on extended TDY is now being re-examined. Major command opinions have been solicited on how to best handle these types of situations. Responses indicate a desire to allow commanders more flexibility in accommodating situations where their people need special testing considerations. These changes will place more responsibility on commanders--they should best know the situation and how it affects their airmen.

Another change to WAPS is the Staff Sergeant (SSgt/E-5) testing and promotion method. All grades, except SSgt, are tested and promoted once a year. SSgts have traditionally been tested and promoted twice a year because of their large numbers. Offsetting this rationale, it is extremely manpower intensive to conduct testing and run promotions for the same group of individuals twice per year. With the reductions in manning of AF personnel offices, we needed to improve efficiency. The system was reviewed and the decision made to eliminate one SSgt testing cycle per year. Although this will create a larger testing requirement within one cycle, manpower and computer resources will ultimately be saved, and a redundant cycle eliminated. Additionally, the requirement to develop and produce two versions of the SSgt test will be eliminated. This alone will shave time off the development effort of the subject-matter experts and save the costs associated with printing and distribution. All required changes have been made and publicized for an annual testing/promotion cycle for SSgts effective in 1992. These changes are expected to save the Air Force approximately 30% in manpower and computer resources. As one might expect, the

transition period will cause a few minor problems, but these are being worked (for example, understanding by the Enlisted Force about how the new method will work, delayed testing and promotion opportunity for some members during the transition).

WAPS has traditionally focused on individual effort and contribution in performance and knowledge. As a result, WAPS testing is predicated on a system of self-initiative which includes individual study and preparation—thus prohibiting any form of group study. This prohibition has increased test compromise rates and presented difficulties in defending its legalities and justifying why it cannot be used for enlisted promotion testing when it is an accepted and effective technique at all levels of education. As a consequence, the Testing Branch has established a working group to look at the issue of group study from top to bottom to determine if any changes to the current blanket prohibition against group study are warranted. Some questions to be answered are: 1) Are we handcuffing our Enlisted Force by prohibiting a basic study technique used pervasively at all levels of education? 2) Are we hindering our people's opportunity to improve job knowledge and broaden themselves professionally? 3) Are we hindering our people's job performance? and 4) Are we indirectly hurting AF efficiency? A game plan has been established with a corresponding timeline. We are still in the research phase—searching literature, discussing, soliciting views from experts and the other Services. Upon completion of this phase, we will make a continuation decision. At that point, the Enlisted Force may be surveyed and implementation issues associated with a policy change will have to be explored. Whatever the outcome of the working group, any change to the promotion system must preserve its integrity, objectivity, and acceptability by the Enlisted Force.

Proficiency tests are critical to mission accomplishment. For this paper, we will specifically discuss the Foreign Language Proficiency Pay (FLPP) Program. FLPP is essential to the AF readiness capability. Language testing is conducted to identify resources proficient in a certain language. AF members are provided monetary incentives to encourage maintenance of their language skills. Once identified as proficient through testing, they are put in a pool for possible duty should a contingency develop where their language skills are required. Due to budget constraints in 1991, changes were made to the FLPP program to scale down the pay and language requirements. At that time, FLPP was being offered to any member who could achieve proficiency through testing. As a result, the number of qualified linguists far exceeded any AF readiness requirements. This problem dictated quick changes. While working these changes, procedures were developed to streamline the testing function to be less manpower intensive. Increased automation of pay will also be accomplished. Language testing was decreased to one cycle per year (vice two) and major changes to the military pay system were formulated to increase automation. These changes will benefit all individuals involved in the administration of the program by decreasing the workload 25% and FLPP expenditures \$3.5M as compared to FY 91. These changes have especially helped AF personnel offices, the key players in the whole process, by not only cutting down the testing required but by more evenly balancing the testing workload throughout the year between promotion testing and language testing. This will boost the accuracy and security of each testing program and benefit the AF member who is paid FLPP through more timely testing and more accurate pay transactions.

More Efficient/New Technology

Training individuals to be officers and/or pilots is a very expensive effort. Selection procedures then become critical to getting the most for our money by predicting successfully those candidates who will complete training and do a good job for the Air Force. To increase our success rate in this arena calls for adding new technology and making our current systems more efficient. To do this, changes are being made at this time and planned for the future.

With fewer pilot authorizations available and less money for training, selection of the very best candidates for pilot training is a must. To increase successfulness of our selection procedures, new technology is being implemented in our testing program later this year. Pilot candidates will be administered a computerized battery of cognitive and psychomotor tests on the Basic Attributes Tester (BAT). The BAT is a computerized test station with control sticks and a keypad so examinees can observe the monitor while simultaneously responding with the control sticks and keypad. The control sticks are used in several batteries to measure psychomotor abilities required for success in pilot training. The Pilot Candidate Selection Method will transmit summary scores over telephone lines to a central processing station at Randolph AFB where they will be weighted with candidates' Air Force Officer Qualification Test (AFOQT) composite scores and flying hours to produce a percentile score. This percentile score will reflect the ranking of each examinee among all previous pilot candidates who have taken the BAT. It is estimated that \$1M a year will be saved through this process.

The AFOQT, our primary officer selection instrument, is also undergoing both policy and procedural changes. Current retest policy hinders our ability to identify individuals most likely to complete an officer commissioning program. Research indicates the first or second administration is the most predictive of success in an officer training program. Current policy requires individuals to retest if their last set of scores is more than 2 years old and limits retesting to twice. This obviously mandates retesting in many instances. For example, an individual has taken the AFOQT twice and the last administration is over 2 years ago. AF policy requires a retest even though the last administration would provide good information for making a selection decision. Policy will change effective this summer to reflect only one retest after the initial administration and levy no time limit on the life of the scores. This will ensure AF selection decisions are made with the most valid data while also saving time and money in reduced test administration and prevent forced retests due to policies.

AFOQT scoring is currently under great change. The current scoring system is based on computer technology from the 1960's and requires several manual intervention steps. We are now updating our automated processing procedures to utilize the increased capabilities of PCs--bringing the whole system in-house. Once the new scoring system is implemented at AFMPC, it will require approximately half the manpower and computer resources as it does currently.

Looking ahead to the future, we realize computerized testing will become increasingly important. We are moving toward this goal by using the BAT to lead the way. With this system already in place, adding the AFOQT to the computer would be a logical next step. We have already planned for this in our budget and are discussing future implementation with our systems experts. We envision

the AFQOT will be administered on a computer, responses transmitted over phone lines, and scored automatically at a central site. The number of machines available will need to be increased to accommodate the number of candidates and the time it takes to administer the test. An analysis will be conducted to determine feasibility and cost effectiveness.

Down the line, as our experience base is built in both military and civilian testing communities, we will be assessing whether computerized promotion and proficiency tests are feasible. The "how to's" of each will need to be determined, but even more importantly, the effect of computerized testing on the equity and fairness of the testing system will have to be thoroughly examined. We will be working hand-in-hand with the research experts in our efforts to pursue an operational testing program of this nature.

The AF Personnel Testing Function will continue to strive for improvement and look for more efficient ways of doing business. We definitely reject the notion of "that's the way we've always done it" as a reason for doing something—yet we will not change just to appear we are using the latest technology. We've made significant progress recently, but still have much to accomplish. Our charter will be to ensure that any and all changes made will benefit all AF personnel and decision-makers while maintaining the integrity and fairness of the testing programs.



Improving Career Selection with the Armed Services Vocational Aptitude

Battery 18/19 Career Exploration Program

92-17661

Kenneth W. Wegner
Boston College

Abstract

The purpose of this study was to develop an improved system for using interest inventory scores in the new ASVAB 18/19 Career Exploration Program. Several procedures for assigning Holland vocational typology codes to occupations were compared across three of the most frequently used classification systems. The results indicate that using rank order by elevation of mean typology scale scores is a highly efficient and consistent method of determining codes. This method appears to provide the most valid approach to improving the differentiation of occupational interest types on the OCCU-FIND chart in the Career Exploration Program.

The new Armed Services Vocational Aptitude Battery (ASVAB) 18/19 Career Exploration Program is scheduled to be available in July 1992 (Wall & Zytowski, 1991). In addition to revised test content and reporting format, a new Military Careers manual and Exploring Careers: The ASVAB Workbook have been developed to assist high school youth and potential enlistees to explore military and civilian careers. One of the new features of Exploring Careers is the inclusion of Holland's (1985) Self-Directed Search (SDS) as a measure of vocational interests. The SDS has four parts measuring competencies, activity and occupational preferences, and estimated abilities. The resulting scores indicate an individual's orientation to six vocational types: Realistic (R), Investigative (I), Artistic (A), Social (S), Enterprising (E), and Conventional (C). Holland recommends the use of the three highest type scores in career exploration.

After completing the SDS and other exercises in Exploring Careers, the three SDS interest codes, two ASVAB aptitude codes, up to four work values, and civilian and military educational information are used to explore 201 occupations on the OCCU-FIND chart. The OCCU-FIND chart utilizes latent image printing to help students match their characteristics to those of various occupations. For example, one may select their most important work values and mark a line with a special pen across all occupations to determine the fit between them. A star appears under each occupation for which the work value is important. One could also draw a line through the column for various occupations to determine which best fit their interest types, aptitude codes, and other characteristics. Although none of the final materials are available to the public at this writing, Wall and Zytowski (1991) have made various conference presentations introducing this new program. This chart lists 201 civilian and military occupations selected from the Occupational Outlook Handbook (Department of Labor,

1990) and Military Career Guide (Department of Defense, 1988). These occupations were first grouped by their Dictionary of Occupational Titles (DOT) (Department of Labor, 1977) classifications. Three letter codes were then assigned using the Dictionary of Holland Occupational Codes (DHOC) (Gottfredson, et al., 1982), and expert judges analyzed the codes to select the primary Holland type for that cluster.

The focus of this study was on the use of Holland's SDS results with the OCCU-FIND chart in selecting military or civilian careers for further exploration. The inclusion of the SDS permits a more precise measurement of interests than was available before. However, using only a primary Holland type to classify an occupation does not provide adequate differentiation between occupations grouped within a particular primary type. For example, Holland (1985) indicates that approximately 40% of high school boys score highest on the SDS Realistic scale, and approximately 50% of occupations in the labor market have a primary classification of Realistic. This suggests that at least a secondary code type is needed for each occupation to narrow the search to those which fit best.

Holland (1985) uses a three letter type code to classify occupations (e.g., a Systems Accountant is coded Conventional, Social, Enterprising or CSE). Holland, 1985, p. 15) states that the assigned codes are approximate, not precise, and the first letter is the most important, descriptive, and reliable. Codes may shift slightly from sample to sample within an occupation, but the shifts most frequently involve a reversal of the second and third letters (e.g., a Cost Account is coded CES). Holland's assignment of codes is based primarily on occupational analysis rather than administering the SDS to occupational criterion groups.

Johannson (1986) utilized the Holland typology in developing general theme scales for his Career Assessment Inventory - Enhanced Version (CAI). Profiles of mean theme scale scores were developed for the 111 combined sex occupational criterion groups of the CAI. The intercorrelations of each theme score with the occupational scale score were used to determine a Holland type code. Thus, a code of C was assigned to Accountant because the Accountant scale had a high positive correlation with the Conventional theme, but no significant correlation with any other theme. This system produced three letter codes for 26% of the occupations and two letter codes for another 37%, but only primary codes for the other 37%.

Hansen and Campbell (1985) used a similar approach to classifying 106 occupations on the Strong-Campbell Interest Inventory (SCII). The SCII has separate sex criterion groups for 101 of the occupational scales. Mean general occupational theme scores and intercorrelations of these scores and the occupational scale scores were used to determine Holland codes. However, the elevation of the mean score on the theme scale was used as the first criterion for assigning a code letter and the intercorrelation as a secondary criterion (as opposed to the Johannson approach above). Thus the Accountant scale on the SCII was assigned a code of C because the only significantly high mean theme score was Conventional. This approach resulted in the assignment of

three letter codes to 21% of the occupations and two letter codes to another 43%, but only primary codes to the other 36%.

The preceding indicates the difficulty in determining a two or three letter Holland typology code for occupations. Of the methods described above, the use of mean theme scale scores to assign Holland typology codes would appear to be the most precise and valid approach. The first purpose of this study was to analyze the degree of agreement between the type codes assigned to like named occupational groups appearing on the CAI, SCII and in Holland's (1985, pp. 56-63) Alphabetic Index to Occupational Classifications and Codes (AIOCC). The second purpose was to determine if a valid method of assigning two letter Holland type codes could be developed for use with OCCU-FIND.

Method

In order to analyze agreement between classification systems, the like-named occupation scales (with at least two Holland type letter codes) were matched across the CAI, SCII, and Holland AIOCC. Of 111 scales on the CAI, 48 matched those in the AIOCC. Of the 106 scales on the SCI, 47 were also found in the AIOCC. The first two letters of the codes for each scale were then compared to determine percent of agreement.

The next step was to use the same procedure to compare Holland codes for like-named occupations that appear on both the CAI and SCII. The purpose of this analysis was to determine level of agreement between the first two letters of codes for occupations on both inventories. Only 30 scales matched by name on both inventories. A possible reason for this relatively low number of overlapping occupational scales is that the SCII is focused more on professional level occupations, while the CAI contains more sub-professional occupations.

The final step in the procedure was to determine which systems of assigning two letter Holland codes to occupations was most efficient and consistent. As indicated previously, the systems used on the CAI and SCII did not permit the assignment of more than a single primary Holland code to over one-third of the occupations on each. Thus if a system could be developed which would add a second code letter for these occupations, it would be useful in assigning two letter codes to occupations on the OCCU-FIND. Since the Holland AIOCC and SCII systems are both based on rank order of most important themes, two letter rank order codes (based on elevation of mean theme scores) were developed and compared to their assigned primary and secondary codes on the CAI and SCII occupational scales.

Results

1. Of the 48 like-named occupations in the Holland AIOCC which also had multiple letter codes on the CAI, only 23% had identical codes through the first and second letters.

2. Of the 47 like-named occupations in the Holland AIOCC which also had multiple letter codes on the SCII, only 36% had identical codes through the first and second letters.

3. Of the 30 like-named occupations with multiple letter codes on both the CAI and SCII, 83% had identical codes through the first and second letters.

4. When the 30 like-named occupations with multiple letter codes on both the CAI and SCII were assigned rank-order codes by elevation of theme score means on each inventory, 90% had identical codes through the first and second letters.

5. When rank-order codes were assigned by elevation of mean theme scores of the 70 CAI occupational scales with multiple letter codes, 79% had identical codes through the first and second letters. Eleven percent of the inconsistency was due to the fact that Johansson's (1986) system did not result in the assignment of a primary code based on the highest mean theme score for eight of the 70 occupations.

6. When rank order codes were assigned by elevation of mean theme scores to the SCII occupational scales with multiple letter codes, 92% had identical codes through the first and second letters.

Discussion

The type codes found in Holland's AIOCC tend to be inconsistent with those developed for like-named scales on the CAI and SCII. The major reasons for this are probably different decision rules and differences in composition of the occupational criterion groups used. Also, Holland (1985, p. 13) states that a range of personalities may be found in any field, and this heterogeneity may contribute to discrepancies in the codes of two like-named occupational samples. There was a significantly higher percentage of agreement on two letter codes between like-named CAI and SCII scales. Also, it appears that using rank order of elevation of means of occupational groups on Holland theme scales would result in consistency level of 90% on the SCII and slightly less on the CAI. These results suggest that ranking the mean theme scores is a viable approach to developing two letter codes for those occupations with single letter codes on the CAI and SCII, and using them to code similar occupations on the OCCU-FIND chart. Since the complete new OCCU-FIND is not available at this writing it is impossible to tell how many of the 201 occupations could be coded using data from Holland theme scales on the CAI, SCII and similar inventories. However, the following illustrates the application of the suggested coding system to sample occupations found in the Military Career Guide (1988).

Realistic (R)

Aircraft Mechanics	RI
Auto Mechanics	RC
Barbers	RE
Firefighters	RS

Investigative (I)

Computer Programmers	IR
Pharmacist	IC
Physician	IA
Nurse	IS

Artistic (A)

Graphic Designers	AI
Interpreters	AS
Reporters	AE
Photographers	AR

Social (S)

Dental Specialists	SC
Chaplain	SA
Trainers	SE
Speech Therapist	SI

Enterprising (E)

Band Manager	EA
Optician	ER
Personnel Specialist	ES
Transportation Specialist	EC

Conventional (C)

Accounting Specialists	CE
Food Service Specialists	CS
Infantrymen	CR
Secretaries	CA

References

Department of Defense (1988). Military career guide. Washington, DC: U.S. Printing Office.

Department of Labor (1977). Dictionary of occupational titles. (4th ed.) Washington, DC: U.S. Printing Office.

Department of Labor (1990). Occupational outlook handbook. Washington, DC: U.S. Printing Office.

Gottfredson, C.D., Holland, J.L., and Ogawa, D.K. (1982). Dictionary of Holland occupational codes. Palo Alto, CA: Consulting Psychologists Press.

Hanson, J.C. and Campbell, D.P. (1985). Manual for theSCII/SVIB. (4th ed.). Palo Alto, CA: Consulting Psychologists Press.

Holland, J.L. (1985). Manual for the Self-directed search. Odessa, FL: Psychological Assessment Resources.

Johannson, C.B. (1986). Manual for the Career Assessment Inventory - Enhanced version. Minneapolis, MN: National Computer Systems.

Wall, J.E. and Zytowski, D. (1991). Linking student characteristics to occupations: the ASVAB career exploration program. Paper presented at the American Association for Counseling and Development Convention, Reno, NV, April.



Military vs. Business Leadership

Suggestions from Career Description Analysis

David D. Van Fleet, Ph.D.
Arizona State University West
Tim O. Peterson, Ph.D.
University of Tulsa

92-17662



Abstract

Comparisons of military and business leadership may be useful if carefully conducted. The data presented in this paper could have been misused to show that military and business leaders were "alike." A more correct comparison is presented and suggestions about future research are made.

In an earlier study on military leaders, Van Fleet and Yukl (1986) found some tentative differences between military and business leaders. However, the comparability of the data used for that work was tenuous. However, in that same monograph, they report the results of using content analysis on career descriptions as a method of learning more about military leaders. A relatively exact replication of that effort was done on a group of business leaders to see what similarities and/or differences might occur when more comparable data sets are employed.

Method

The only difference in the methods used was a slight difference in the categories used -- in the earlier military study, role clarification and goal setting were combined for a total of 22 categories while in the replication (business study) they were separate for a total of 23 categories. However, in each study graduate students were trained to do content analyses using the definitions of the behavior categories in Yukl's taxonomy of managerial leader behaviors. Each student read a career description of his or her choice (either an autobiography or a biography). As he or she read the book, tally marks were made on a worksheet each time a behavioral description was encountered. Because the military career descriptions tend to emphasize heavily wartime experiences, a group of business leaders was selected who had lead their organizations through turbulent times, particularly its founding and/or major transformational periods. The business leaders, then, could be thought of as representing entrepreneurial leaders.

Results

As shown in Table 1, the tallies were converted into frequencies for comparison and a significance test run across all leaders on the averages weighted by the number of incidents in each career description just as had been done in the military study. Four categories of leader behavior occurred with a frequency greater than that which would have been expected based on chance alone. They were planning, innovating, goal setting, and emphasizing performance.

Table 1

Content of Career Descriptions of Business Leaders

Behavior Category	Executive (percent frequency of identifiable occurrence of behavior in book)															Unweighted Weighted average
	Alllyn	Ash	DeLoorean	Fatjo	Ford	Geneen	Iacocca	Kamiya	Kennedy	Marcus	Munn	Watson	Wood	Sloan	Watson	
Planning	5	2	9*	13*	12*	3	4	14*	25*	7*	4	2	10*	11*	9	9*
Innovating	14*	5	9*	7*	12*	5	15*	7	14*	6	6	2	8*	5	8	8*
Goal Setting	1	6	5	21*	3	4	4	7	0	4	4	8*	6	6	6	8*
Emphasizing Performance	3	3	8*	6	8*	15*	6	3	0	12*	6	4	6	2	6	7*
Showing Consideration	9*	11*	3	4	5	8*	3	6	3	10*	13*	7	7*	6	7	6
Problem Solving	4	4	8*	9*	4	7*	15*	7	12*	3	9	0	4	6	7	6
Facilitating Cooperation and teamwork	8*	3	5	8*	5	14*	6	7	0	1	2	8*	4	6	6	6
Monitoring the environment	3	4	7	3	3	3	8*	10*	2	8*	4	5	10*	7	6	6
Inspiring Subordinates	5	12*	3	4	6	6	3	5	5	5	6	11*	6	2	6	5
Monitoring Operations	5	1	4	2	4	12*	1	4	2	5	2	4	3	5	4	4
Facilitating the Work	3	0	8*	9*	2	0	4	4	0	4	2	5	2	8*	4	4
Structuring Reward Contingencies	5	4	2	0	8*	3	2	0	11*	0	11*	7	3	2	4	3
Training-Coaching	8*	9	2	0	9	0	0	3	0	5	0	7	2	0	3	3
Clarifying Work Roles	3	4	2	3	5	3	1	1	2	1	2	5	5	7	3	3
Representing the Unit	8*	0	1	1	0	2	5	7	0	5	4	7	3	1	3	3
Encouraging Decision Participation	0	6	4	4	3	2	3	2	0	2	8	0	4	5	3	3
Delegating	1	4	5	1	2	3	4	0	6	11*	2	0	7*	7	4	2
Providing Praise and Recognition	1	9	0	0	3	2	4	2	6	1	2	8*	1	0	3	2
Career Counseling	0	5	1	2	2	1	7	5	8	3	2	0	3	0	3	2
Disseminating Information	6	1	6	2	0	2	1	1	0	2	2	4	1	7	2	2
Criticizing	2	5	6	1	1	1	2	1	2	1	4	2	1	5	2	2
Managing Conflict	3	0	2	1	3	1	1	6	3	2	6	0	2	1	2	2
Administering Discipline	3	2	2	0	2	2	1	0	0	1	2	5	2	0	2	1
Number of Incidents	153	60	128	480	129	264	139	117	64	360	53	168	267	131		2,513
Autobiography (A) or Biography (B)	A	A	A	A	A	A	A	A	A	A	A	B	B	A		

*p<0.05 (one-tail z test for proportions)

Executive and Major Corporate Identifier:

Stanley C. Alllyn NCR	Tom J. Fatjo, Jr. Browning Ferris Industries	Lee Iacocca Ford; Chrysler	Stanley Marcus Neiman Marcus	Thomas J. Watson, Sr. IBM
Mary Kay Ash Mary Kay Cosmetics	Henry Ford Ford Motor Company	Shoter Kamiya Toyota	Henry Lightfoot Munn Munn-Bush Shoe Company	Robert E. Wood Army General; Sears
John Z. DeLoorean GM; DeLoorean Motor Co.	Harold Geneen ITT	Gerald S. Kennedy General Mills	Alfred P. Sloan General Motors	

The earlier study of military leader obtained the results shown in Table 2. Two or more Correlations between the two results must be made because of the slight differences in the number of categories used. One must drop either goal setting or clarifying work roles from the business study. As it turns out, it makes no difference in terms of interpreting the results. All tests of the hypotheses of no difference in weighted averages across the two samples were significant (+0.52 dropping goal setting and +0.67 dropping clarifying work roles, both significant at the .01 level) or in rankings (+0.54 dropping goal setting and +0.65 dropping clarifying work roles, both significant at the .001 level). That analysis, then, would suggest that these business and military leaders are relatively similar. However, such a conclusion would be in error.

Table 2

Results of Content Analysis of Career Descriptions
of Military Leaders (Van Fleet & Yukl, 1986)

Significantly greater than chance amount

clarifying work roles and objectives (goal setting)
monitoring the environment
planning
inspiration

Not significantly greater than chance amount

performance emphasis
monitoring operations
problem solving
representing the unit
facilitating cooperation and teamwork
autonomy-delegation
information dissemination
innovation
providing praise and recognition
facilitating the work
showing consideration
managing conflict
encouraging decision participation
training/coaching
constructive criticism
administering discipline
structuring reward contingencies
career counseling

In each of the studies, only four of the 22 or 23 categories of behaviors were at an amount significantly different from what one would expect under a random distribution of incidents. Thus, those are really the only ones which should be compared. Planning is in both lists as is goal setting, although in the military study it is mixed in with clarifying work roles. Emphasizing performance was significant in the business study and it barely missed being significant in the military one. However, inspiration was significant in the military but not in the business study, and innovating was significant in the business but not the military study.

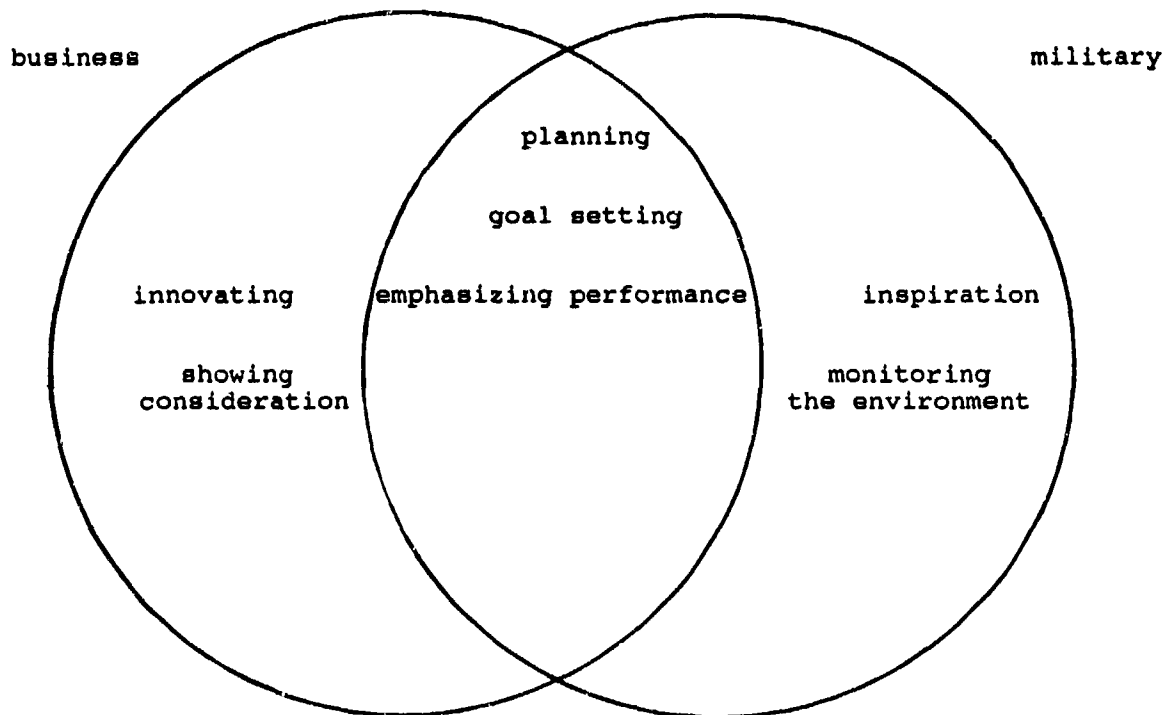
Conclusions

There may be many similarities between business and military leaders as exemplified by the correlations between rankings of weighted frequencies of

categories of behavior. However, dissimilarities are more likely to be the factors most closely with effectiveness in the two different contexts. In this particular instance, the career descriptions suggest, as shown in Figure 1, that planning, goal setting, and perhaps performance emphasis are important in both contexts while innovation and showing consideration seem more pertinent to business entrepreneurs while inspiration and monitoring the environment seem more pertinent to military leaders.

Figure 1

Significant Behavior Categories For Two Studies



Great care needs to be taken with these conclusions, however. Biographies and autobiographies can easily reflect personal biases and selective memory than actual behaviors which occurred. Further, each of these samples is limited -- the business sample is more of an entrepreneurial sample than one of business managers or executives in general and the military sample is heavily weighted toward upper ranks and war-time conditions. Nevertheless, this sort of analysis is suggestive and, if confirmed through the use of other methodologies and other samples, could be important to those involved with leadership education in the military.

References

- Allyn, Stanley C. (1967). My Half Century with NCR. New York: McGraw-Hill.
- Ash, Mary Kay (1984). On People Management. New York: Warner Books.
- DeLorean, John Z. (1979). On a Clear Day, You Can See General Motors. New York: Avon Books.
- Fatjo, Tom J., Jr., & Miller, Keith (1983). With No Fear of Failure. New York: Barkley Books.
- Ford, Henry (1922). My Life and Work. New York: Doubleday.
- Geneen, Harold, with Moscow, Alvin (1984). Managing. New York: Doubleday.
- Iacocca, Lee (1984). Iacocca: An Autobiography. New York: Bantam Books.
- Kamiya, Shotaro (1976). My Life with Toyota. Tokyo: Toyota Motor Sales Co., Ltd.
- Kennedy, Gerald S. (1971). Minutes and Moments. Minneapolis, Minnesota: Hart Press.
- Marcus, Stanley (1974). Minding the Store -- A Memoir. Toronto, Ontario: Little, Brown.
- Nunn, Henry Lightfoot (1953). The Whole Man Comes to Work. New York: Harper & Row.
- Rogers, William (1969). Think -- A Biography of the Watsons and IBM. New York: Stein and Day.
- Sloan, Alfred P. (1964). My Years with General Motors. New York: Doubleday.
- Van Fleet, D.D., & Yukl, G.A. (1986). Military Leadership: An Organizational Behavior Perspective. Greenwich, Connecticut: JAI Press.
- Worthy, James C. (1984). Shaping an American Institution: Robert E. Wood and Sears Roebuck. Urbana, Illinois: University of Illinois Press.



AN EXPLORATION OF TRANSFORMATIONAL LEADERSHIP AND ITS EFFECTS

92-17663



Ruediger Mueller
University of Wisconsin, Oshkosh
Marshall Sashkin
Office of Educational Research & Improvement
U.S. Department of Education
William E. Rosenbach
Gettysburg College

Abstract

This study extends the authors' 1988 analyses of the Leadership Description Questionnaire by reporting a principal components analysis of fire service executives' self-report data and data obtained from their subordinates' assessment of their leadership style. Multiple regression is then used to predict satisfaction, effectiveness and organizational commitment using transactional and transformational leadership behaviors. Results are discussed in terms of the appropriateness of the instrument as a measurement tool.

In 1988, Mueller and Rosenbach reported the results of research which dealt with principal components analyses of the Leader Description Questionnaire (LDQ). The LDQ is an abbreviated and somewhat revised version of the Multifactor Leadership Questionnaire (MLQ), (Bass, 1985, Clover, 1986). Mueller and Rosenbach's analyses of 1304 subordinate ratings of 171 leaders from three different populations (business, fire service, and church) showed that no substantial multicollinearity existed among the questionnaire items, thus permitting the use of regression analyses to predict follower satisfaction and leader effectiveness. In the second part of their study they were unable to find conclusive evidence of the presence of specific transactional and transformational leadership behaviors and hence suggested further research.

The current study extends our original analyses in two directions. First, a principal components factor analysis is repeated, but with the addition of a self-report data set to the data obtained from subordinate assessments of managers. The resultant factor structures are then compared. Second, multiple regression is used to predict satisfaction, effectiveness, organizational commitment, and role modeling, using transactional and transformational leadership behaviors. Subordinate reports of leaders' behaviors are used to predict leaders' ratings of satisfaction, effectiveness, organizational commitment, and role modeling. Similarly, leaders' self-reports are used to predict subordinates' reports of satisfaction, effectiveness, commitment, and role modeling.

Measures

Several years ago we received permission from Bass to revise and modify the MLQ

for use in research we wished to conduct. Since then we have conducted a number of psychometric analyses and made several revisions of our modified measure, the LDQ. While the LDQ retains many similarities with the MLQ, the latter has undergone many revisions and modifications itself, and it would be misleading to suggest that the two instruments are equivalent measures, even though both aim to assess several dimensions of transformational and transactional leadership.

Over the last few years the LDQ has been used in the context of leadership training seminars in a variety of different settings, both to develop a database and to conduct research on leadership training effects and outcomes. The largest current database incorporates LDQ data from seminars given for the National Fire Academy, a Federally-sponsored organization chartered to develop effective leadership among professional and volunteer fire service executives from all over the United States and Canada. Over 130 individuals assessed themselves and were assessed by an average of about seven subordinates, using the LDQ; the total number of subordinate reports is over 1400 observations. The self-report and subordinate-report databases permit us to ask two research questions:

Does the LDQ measure a set of distinct categories (factors) of leadership behavior, as intended?

Are those leadership behaviors predictors of subordinate satisfaction, effectiveness, commitment, and role modeling reports?

The first question is relevant if the LDQ is to be used for planning purposes, for example, to decide about management development programs. The LDQ measures three transactional leadership behaviors: Contingent Reward (CR), Laissez-faire (LF), and Management by Exception (MBE). CR involves granting rewards for performance while LF is essentially an absence of leadership. MBE means taking management action only in exceptional situations. The five transformational leadership behaviors measured are: Charisma (CH), Individualized Consideration (IC), Intellectual Stimulation (IS), Inspiration (I), and Motivation Beyond Expectations (MBEX). The LDQ also includes several "outcome" measures: Organizational Commitment (OC), Role Modeling (RM), Satisfaction with Leader (S1), Satisfaction with Leadership (S2), and Effectiveness (E). It is important to note that the effectiveness measures are, like all of the others, based on perceptions and not on objective indicators; effectiveness, as measured here, must be treated with caution.

Results: Factor Analyses

Complete correspondence of the factors obtained from self-report data and factors derived from subordinate-report data was not achieved, but a number of factors from the two analyses do correspond quite closely. Neither do the factors perfectly correspond to the eight scales that comprise the LDQ though, again, the results are encouraging. Two factors did correspond exactly for the self and subordinate data. The first combines the two dimensions of Inspiration and Motivation Beyond Expectation. This is intuitively appealing, as a high level of motivation could be expected to result from inspirational

leadership. The second pair of corresponding factors consisted of items assessing Intellectual Stimulation. A third factor appeared consistently in both datasets, composed of the three items we created to tap organizational commitment (items not part of the original MLQ).

Several additional factors, while not as closely corresponding, were still close enough to lend support to the construct validity of the instrument. MBE questions all loaded on the same factor, but for self-report data only. Results for subordinate-report data were actually quite similar, except that not all of the MBE items formed a single factor.

Despite some encouraging factor clarity and consistency across data sets, the transactional and transformational behaviors do not seem to be as distinct in practice (as seen by subordinates and leaders) as theory would seem to suggest. In both datasets, for example, there is a factor composed of the same two items, one relating to CR (a transactional leadership behavior) and the other tapping IC (a transformational behavior). (The factor derived from subordinate-report, however, includes a second IC item, thus appearing to more strongly represent a transformational behavior category.) Four CH-related items load on the same factor for subordinate-report data (along with one LF question). Based on the self-report data, however, CH forms two factors. Apparently leaders see two aspects to charisma.

Results: Regression Analyses

We created matched pairs of data records in order to perform these analyses. We paired the leader's self-assessment data with the means of subordinate-report data. We then performed two separate sets of regression analyses. In both the dependent variables were Satisfaction, Commitment, Role Modeling, and Effectiveness while independent variables were the leadership behaviors as measured by the eight LDQ scales. In the first set of analyses, however, the dependent variables were obtained from subordinate ratings, while independent variable data were taken from leaders' self-ratings. For the second set of analyses, dependent variables were leaders' ratings of outcomes (subordinate satisfaction with the leader, effectiveness, etc.) while the independent variable measures were taken from subordinate-report data. In this manner we avoided the problem of same-source data bias.

The predictive ability of all of the models was very strong, as shown by the very high coefficients of multiple correlation. All of these results were highly significant ($p < .001$). A comparison of the two sets of analyses shows that there is no discernable pattern in reference to self or subordinate assessment data. That is, no pattern exists that would permit rating either the subordinate or the self-assessment data as the superior predictors. However, the origin of the data must be taken into account. These data were collected in the context of training programs conducted by the National Fire Academy. Neither leaders nor subordinates had any reason to feel threatened by the results; indeed, both could expect improved relationships based on the use of accurate information in the training seminars. Everyone was also explicitly promised complete confidentiality. Our results might be different in a different context.

Of considerable interest is the fact that regression models using **transactional** leadership behaviors as predictors perform as well as those using **transformational** behaviors. This may underscore the notion that in today's organizations both transactional and transformational leadership are needed.

The coefficient of multiple correlation is only one measure of the predictive power of a regression model. One would also want to look at the individual predictor variables and their influence on the model. It is here that the analyses reveal some problems. The importance of individual predictor variables varies greatly between models and some variables are not individually significant. This suggests the presence of interaction effects, hardly surprising given the previously reported factor analysis results. The predictor variables used in the two sets of regression analyses are the means of the sums of individual item scores for items that make up one or another of the eight LDQ scales. In the absence of interaction effects the factor analyses would have resulted in one factor each for each of the eight dimensions, with items loading strongly only on a single factor. Instead, the factor analysis results showed factors that combined items from different scales. Moreover, some items loaded high on more than one factor. Therefore, while transactional and transformational leadership behaviors are good predictors of the outcomes defined here - satisfaction, commitment, effectiveness, and role modeling - it is not possible to isolate specific dimensions that are better predictors than others. Further analysis and experimentation will be needed to identify more clearly which of the eight dimensions are most important in predicting various outcomes.

Discussion and Conclusion

Our factor analyses show a high though not perfect degree of consistency between subordinate and self-assessment data. Furthermore, many of the LDQ scales seem to have a reasonably high degree of internal consistency. However, it was not possible to identify eight factors that would relate precisely to the three transactional and five transformational leadership behavior dimensions. While some (like MBE) are indeed found for one or both of the datasets, others (e.g., I and MBEX) were combined in a single factor, and still others (such as LF items) did not load onto one single factor. Therefore, while the results of the analyses reported here are encouraging, further fine tuning of the LDQ items and scales is indicated. Indeed, Bass has revised the MLQ many times since producing the version that was the original source of the LDQ.

Despite the encouraging results of the regression analyses, we must note that the coefficients of multiple correlation might be much lower in other contexts. When data are collected in more realistic, that is, potentially threatening, situations, in which respondents are motivated to make less than fully truthful reports, the same results would probably not be obtained. For example, if compensation and promotion are thought to depend on one's relationship with one's boss, we could expect less forthcoming and honest assessment responses from subordinates, leading to lower multiple r 's.

We must also note that effectiveness was reported by the respondents, not assessed by means of objective measures. In an organizational setting, objective measures would be far more desirable. Furthermore, time series analysis of data might show deterioration or improvement over time and could, then, prove even more useful for making appropriate interventions.

Despite these limitations, we have shown that the LDQ, derived from Bass' MLQ and based on his concepts of transformational leadership, holds up relatively well, both in terms of construct and concurrent validity analyses. With some revisions in the instrument, we expect to continue to study transformational leadership using the LDQ as one important measurement tool.

References

- Bass, B.M. (1985) Leadership and Performance beyond Expectations. New York: The Free Press.
- Clover, W.H. (1986) Item Reduction of the Multifactor Leadership Questionnaire. U.S. Air Force Academy Working Paper.
- Mueller, R.; Rosenbach, W.E. (1988) Principal Components Analysis for Multicollinearity and Homogeneity of Transactional and Transformational Leadership Dimensions. Proceedings of the Eleventh Symposium on Psychology in the Department of Defense.
- Rosenbach, W.E. & Mueller, R. (1989) Transformational and Transactional Leadership of Business, Church, and Fire Service Executives. Advances in Industrial Organizational Psychology. Fallon, B.J., Pfister, H.P. & Brebner, J. (eds.) Amsterdam: Elsevier.

Note: A complete copy of the paper, including summary tables of the factor and regression analyses is available from the authors on request.



**Centrality of Work: The General Moderator Hypothesis and
Scale Development**

Frederick W. Gibson
Kevin A. Ashley
Douglas R. Lindsay
United States Air Force Academy

92-17664



Abstract

Based on Locke's (1976) Value Theory, centrality of work is proposed as a moderator of several relationships between work-related variables, such as that between job satisfaction and performance. As a first step in examining this hypotheses, we developed a prototype Centrality of Work (COW) Scale. A sample of 135 Air Force Academy faculty and cadets completed the prototype COW Scale. The data were used to examine the factor structure of the items, and to assess differences between the subgroups.

One of the classic paradoxes in Industrial/Organizational psychology is that several intuitive relationships between work variables are not supported in the data. Possibly the most perplexing of these is the relationship between job performance and job satisfaction. Time and again, research has shown the correlation between these two variables to be insignificant. Brayfield and Crockett (1955) determined that "no appreciable relationship" could be found between the two variables, and Locke (1976) came to the conclusion that "job satisfaction has no direct effect on productivity" (p. 1334). More recently, Iaffoldano and Muchinsky's (1985) meta-analysis on the satisfaction-performance relationship found an average correlation of $r=.14$. These findings are much weaker than one would expect, and in fact are plausibly viewed as trivial.

Several explanations have been offered for these disappointing results, including more complex causal models, and a reversal of the proposed causal relationship between the variables. However, neither has been very successful in generating supportive data. One promising alternative may lie in identifying subgroups for whom the relationship between, e.g., satisfaction and performance will hold more strongly than for others. In other words, we believe a work-related psychological variable moderates relationships like this. The goal of the present research is to examine the nature of this potential general moderator, which we term centrality of work (COW).

Locke's (1976) Value Theory hypothesizes that the importance individuals attach to a certain value directly impacts the range of affect that value can produce. So, for example, individuals who place great importance on workplace temperature would report greater ranges in satisfaction associated with deviations in temperature than would individuals who place less importance on temperature. The statistical implications are crucial, because the group which does not place great importance on temperature in effect exhibits a restricted range of satisfaction. For that group, the correlation between satisfaction and temperature (or satisfaction and any variable of investigation) would be attenuated. In the context of our research, the importance one attaches to his or her job (centrality of work) determines the range of satisfaction which job

characteristics might produce. Low centrality individuals would exhibit a restricted range of satisfaction since they are psychologically "disengaged" from the workplace; more important, their restricted range of satisfaction will attenuate the correlation between this variable and others. In sum, the low correlations between satisfaction and performance, for example, may be due to a failure to effectively classify members into subgroups based on work centrality. Thus, if we can identify high centrality groups, we may be more effective in influencing their job satisfaction by manipulating their work environments, and in discovering the relationships between satisfaction and other constructs.

The first step in this research is constructing an effective COW scale. Several measures of this or related constructs are available; however, few of them incorporate centrality as a multidimensional construct. Others may be criticized for their formats, scoring procedures, and/or factor makeup. We therefore decided to build a relatively comprehensive item pool using previous measures wherever available, and to investigate the factor structure of the resulting instrument.

Method

Construction of the COW Scale

An initial item pool was developed from three sources: the authors; a survey of faculty members; and a variety of instruments addressing work orientation and job involvement. In the last category, scales by Dubin (1963), DuBrin (1984), and Lodahl and Kejner (1965) all contributed items. The Lodahl and Kejner (1965) scale was of particular interest, since it listed items sorted into several easily interpretable factors; indifference/hopelessness; job involvement; duty; and guilt/avoidance. One purpose of this study was to compare the factor structure of the COW scale with that reported by those authors.

This initial item pool was administered to 3 cadets and 10 faculty members in the department of Behavioral Sciences and Leadership as a pre-test to solicit feedback as to which items tapped centrality of work. The 77 which remained were used in the final Centrality of Work scale. The COW employed a five-point Likert scale, where "1" indicated disagreement with the statement and "5" indicated agreement. Fifteen demographics questions were added to the instrument to complete the scale.

Procedure

The subjects consisted of 215 cadets from 2 cadet squadrons, and 56 faculty members from two departments. COW scales were distributed to each member of the sample, and respondents were asked to complete and return them within two weeks. Of the returned surveys, 88 cadet and 47 faculty booklets were usable, for effective return rates of 41% and 84%, respectively. These were combined into one sample (50% return rate) for the purpose of data analysis, but were first coded for group membership.

Results

To determine the factor structure of the scale, the 77 scale items were subjected to principal components and unweighted least squares factor analyses. Both extraction methods yielded similar results, but the unweighted least squares solution was most interpretable and is reported here. Eigenvalues of 1.00 or higher were computed for 22 factors, but a scree test

was used to select 6 factors for further investigation. A quartimax rotation was performed on all the factors.

The dimensions which emerged were identified as: Attachment, positive affect toward work as an activity; Involvement, reported importance of work activities and outcomes; Perfectionism, reported feelings of disappointment accompanying work failures; Spillover, the extent to which work and thoughts of work intruded into other activities; Family History, reported degree to which the respondent's mother and father were "workaholics"; and Instrumentality, reported extent to which work was viewed as merely an instrumental activity. To aid in interpretation, the items which loaded most strongly on each factor are listed below.

Table 1

Representative items of the COW Factors

Attachment

- "I look forward to going to work."
- "Quite often I feel like staying home from work instead of coming in." (R)
- "I look forward to retirement." (R)
- "I would like to retire as soon as I can." (R)
- "I would probably keep working somewhere even if I didn't need the money."

Involvement

- "The biggest thrills I get in life are from my work."
- "My job is the most important thing in my life."
- "The most important things that happen to me involve my work."
- "The major satisfaction in my life comes from my job."
- "I live, eat, and breathe my job."

Perfectionism

- "I feel depressed when I fail at something connected with my job."
- "I'd like to kick myself when I make mistakes in my work."
- "I think of myself in terms of the type of job that I do."
- "I'm really a perfectionist about my work."

Spillover

- "I often bring work home with me."
- "I often stay at work late to complete projects or tasks."
- "I often think about work when I am at home."
- "I am often the first to arrive and the last to leave at work."
- "I am likely to sacrifice family time for work related activities."

Family History

- "My mother was a workaholic."
- "My father was a workaholic."

Instrumentality

- "I work primarily so I can pay my bills and enjoy a few luxuries."
- "Work is meaningless unless it helps you provide for people close to you."

NOTE: (R) denotes items which are negatively loaded on a factor.

To investigate subsample differences and roughly assess the sensitivity of the instrument, we computed "approximate factor scores" by averaging the unit-weighted scores for the items loading most strongly on each factor. The results are listed below. The data indicate that cadets and faculty score highest on perfectionism and spillover, and, surprisingly, score lowest on involvement.

Table 2

Approximate factor score means for subgroups

	Factor					
	Attachment	Involvement	Perfectionism	Spillover	Fam.History	Instrumentality
Cadet	2.73	2.26	3.63	3.69	3.32	3.13
Faculty	3.35	1.90	3.77	3.50	2.72	2.94

We calculated t-tests assessing differences between faculty and cadets on the factors. Results indicated that cadets scored higher on Involvement ($t=2.33$, $p=.02$, $df=128$) and Family History ($t=3.19$, $p=.002$, $df=132$). Faculty scored higher on Attachment ($t=5.07$, $p<.001$, $df=132$). No differences were noted for the Perfectionism ($t=.98$) or Instrumentality ($t=1.09$) factors. We view these data as marginally supporting the sensitivity of the scale.

Discussion

As asserted by Dubin (1963) and Lodahl and Kejner (1965), work centrality is a multidimensional construct. The current study supports such a general view, but the factor structure reported here is different than those reported in earlier studies. In particular, the "indifference" and "guilt" dimensions reported by Lodahl and Kejner were not replicated in this sample; however, their "involvement" factor parallels our dimension of the same name, and "pride" was roughly approximated by Attachment. Moreover, several other dimensions emerged; of these, Perfectionism seems most interesting, since the current Instrumentality (Kidron, 1978) and Spillover (Wilensky, 1960) have already been described as elements of job centrality, albeit not part of any coherent factor structure. On the whole, then, this study marks the first indication that work centrality is more complex than the 4-factor constructs reported heretofore. Obviously, this multivariate makeup must be considered when assessing COW's workings as a potential moderator variable.

Whether these differences in factor structure reflect differences in the item pools of the various researchers or meaningful differences in the samples remains to be seen. Although we are inclined to agree that USAFA cadets and faculty are somewhat perfectionist, we doubt they are uniquely so. However, we view the emergence of this factor as a unique contribution of this data set to theorizing about the centrality construct. Nevertheless, this factor solution needs to be replicated with substantially different samples.

The results of the subsample t-tests indicate the COW scale possesses some sensitivity to group differences. This is particularly heartening in light of the fact that the groups were relatively homogeneous. Moreover, it is interesting to note that both USAFA faculty and cadets reported high perfectionism and spillover scores, but low involvement. One might interpret these data to mean that these groups hold high standards of any work they do, and that they sacrifice other aspects of their lives to do well, but that they do not view their particular jobs as important. This is puzzling, but may be explained in light of the current instability in the military, especially for careers. Nevertheless, until more normative data are available, these conclusions are highly speculative.

We view the future research agenda as consisting of three phases. First, the COW scale needs to be revised based on psychometric findings. We hesitate to trim items at this point since the somewhat unique factor structure produced by this sample needs to be replicated. Second, normative data need to be gathered from different job types and samples representing different demographics. Third, the moderating properties of the construct, as hypothesized in our literature review, need to be investigated. Our hypothesis at this early stage is that Involvement, Spillover, and Instrumentality are the most promising moderators, since they represent the most archetypal examples of psychological investment or engagement in work activities.

References

- Brayfield, A.H., & Crockett, W.H. (1955). Employee attitudes and employee performance. Psychological Bulletin, 52, 396-424.
- Dubin, R. (1963). Industrial workers' worlds: A study of the central life interests of industrial workers. In E. Smigel (Ed.), Work and Leisure. New Haven: College and University Press.
- DuBrin, A.J. (1984). Foundations of Organizational Behavior: An Applied Perspective. Englewood Cliffs, NJ: Prentice Hall.
- Iaffaldano, M.T., & Muchinsky, P.M. (1985). Job Satisfaction and job performance: A meta-analysis. Psychological Bulletin, 97, 251-273.
- Kidron, A. (1978). Work Values and Organizational Commitment. Academy of Management Journal, 21, 239-247.
- Locke, E.A. (1976). The nature and causes of job satisfaction. In M.D. Dunnette (Ed.), Handbook of Industrial and Organizational Psychology. Chicago: Rand McNally.
- Lodahl, T.M., & Kejner, M. (1965). The definition and measurement of job involvement. Journal of Applied Psychology, 49, 24-33.
- Wilensky, H.L. (1960). Work, careers, and social integration. International Social Science Journal, 12, 543-560.



Do Situational Variables Affect Adherence to

Formal Organizational Policies?

Gordon J. Curphy, PhD.

Frederick W. Gibson, PhD.

Gary Macomber, PhD.

United States Air Force Academy

Callie Calhoun, B.S.

United States Air Force Occupational Measurement Squadron

Abstract

A number of researchers have examined how individual differences and situational variables contribute to counterproductive behaviors (Blasi, 1980; Malinkowski & Smith, 1985; Sackett, Burris, & Callahan, 1989), but little is known about the variables related to the reporting of these behaviors. This study used the scenario-based approach described by Motowidlo, Dunnette, and Carter (1990) to determine whether situational variables affect subjects' intended reporting of various organizational policy infractions. Approximately 400 cadets from the United States Air Force Academy completed an instrument made up of 12 scenarios which systematically varied (a) the emotional closeness of the infraction, (b) the severity of the infraction, and (c) the presence or absence of other witnesses to the infraction. The results indicated that all three situational factors played important roles in determining reporting intentions.

Given the number of journal articles and books devoted to the subject, the study of organizational ethics has been a topic of increasing importance over the past decade. Some of this research has been concerned with the relationships among situational and individual difference variables and unethical employee behavior. One important finding in this regard has been that formal ethics policies tend to reduce the incidence of unethical behaviors (Hegarty & Sims, 1979; Trevino & Youngblood, 1990). Although many colleges and universities maintain formal ethics policies addressing cheating, a substantial percentage of students anonymously admit to cheating on a regular basis (Gordon, 1990). Gordon (1990) maintained that the pursuit of grades and fraternity/sorority loyalties are often cited reasons for cheating, however, weak enforcement of ethics policies may also play a crucial role. Although the situational, individual difference, and demographic factors related to the manifestation of unethical behaviors has been investigated, little is known about those factors related to the toleration or reporting of these acts. As such, this study is different from most studies of ethical behavior, as it explores the situational variables related to the reporting of organizational policy infractions rather than the overt performance of these behaviors.

Method

Setting

The United States Air Force Academy (USAFA) is a four-year undergraduate military institution whose mission is to "provide instruction and experience to all cadets so that they graduate with the knowledge, character, and motivation essential to leadership as career officers in the United States Air Force" (Kochanski, 1990, p. 132). In addition to academic and athletic training, cadets receive extensive training and indoctrination in the organizational policies governing cadet behavior throughout their four-year tenure. Several aspects of these policies governing cadet behavior make USAFA unique among colleges and universities. First, any cadet witnessing a policy infraction by another cadet must either report the incident or confront the violator and give him or her the opportunity to self report. Cadets who fail to report another cadet's transgression are guilty of toleration, which itself is an organizational policy infraction. Second, the cadets themselves are responsible for the enforcement and administration of the

policies governing cadet behavior. Cadets investigate, judge, and have a role in determining the punishment for all cadets found in violation of organizational policies.

Measure

This study used written scenarios to assess intent to report various organization policy violations. Three situational variables were varied across the scenarios, which were (a) the seriousness of the infraction ("severity"), (b) the emotional closeness between the transgressor and the observer/reporter ("closeness"), and (c) the presence or absence of other witnesses ("turn-in"). Because of the policies governing cadet behavior, the presence of others should increase the likelihood that the observer would be reported for tolerating the transgression if he or she did not report the incident. Through the use of scaling exercises, several exemplars were chosen which represented three levels of severity, two levels of emotional closeness, and two levels of turn-in. These exemplars were systematically matched to create a completely crossed set of 12 scenarios.

Although the development of the scenarios played a crucial role in this study, the development of a response scale that would be appropriate for all of the scenarios was also important. The first step in the response scale development was the generation of nine statements which spanned the range of possible responses to organizational policy infractions. Eighty cadet participants rated each statement on the relative strength of the response by writing the letter corresponding to each statement on a visual scale from 1 to 25, where higher numbers indicated greater strength of response (i.e., greater intent to report the violator for the infraction). The results of this initial scaling were then used to select five statements which both spanned the response strength continuum and had relatively high levels of agreement (i.e., low standard deviations).

In addition to the scenarios and response scale, several demographic items were included in the final version of the questionnaire. These demographic items assessed cadets' Grade Point Average (GPA), Military Performance Average or MPA (a composite of military ratings from peers, cadet and officer supervisors, academic instructors, and athletic coaches), policy violation history, completion of a philosophy course on ethics, participation in intercollegiate athletic teams or clubs, reported religiosity, class year, and participation in specialized ethics code training.

Administration and Sample

Two versions of the questionnaire were produced by randomly varying the order of the scenarios. Cadets were asked to indicate their responses to all questions and scenarios on an optically scannable form. This questionnaire was administered to four cadet squadrons of approximately 110 cadets each over a three day period in October 1990. Two of the squadrons were chosen because their freshmen had participated in a specialized ethics code training program. The other two squadrons were chosen for convenience; however, because cadets are assigned randomly to squadrons, the overall sample of cadets was believed to be representative of the cadet population.

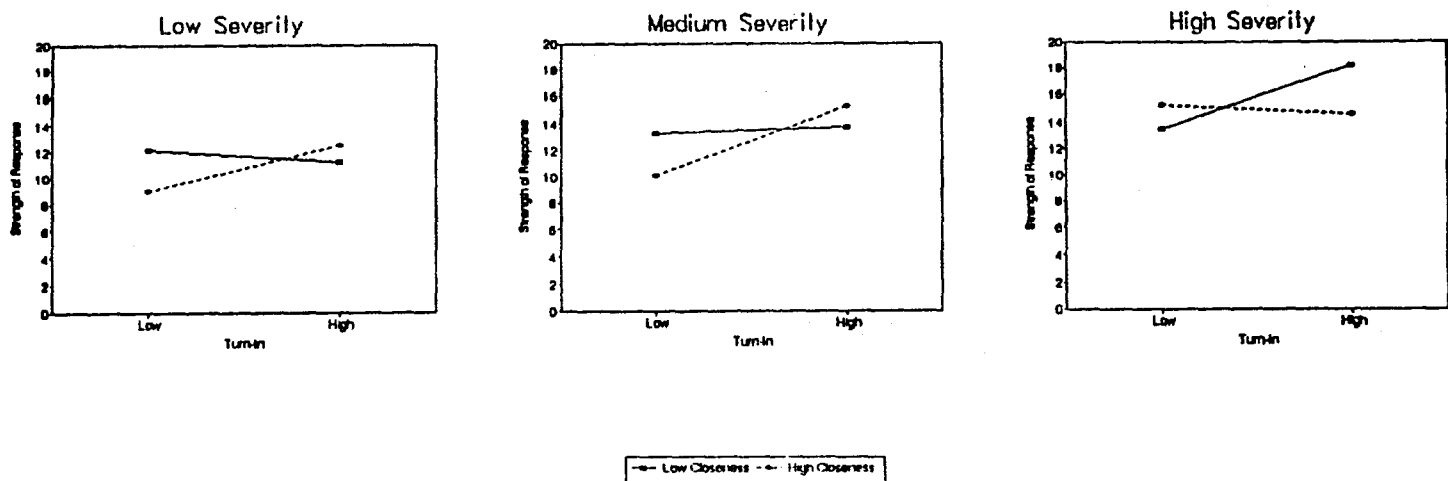
Results

As described in the previous section, this study was planned as a $3 \times 2 \times 2$ repeated measures design, with 3 levels of severity of offenses, 2 levels of emotional closeness, and 2 levels of turn-in. To analyze the data for the effects of these situational factors, we used the repeated measures option in the SPSS-X MANOVA program (SPSS, Inc., 1988). All three main effects were significant in the predicted direction (severity: $F(2, 728) = 148.72, p < .000$), closeness: $F(1, 364) = 36.76, p < .000$, and turn-in: $F(1, 364) = 186.23, p < .000$). Cadets were more likely to report relatively severe offenses and were also more tolerant of offenses committed by close friends than those committed by relative strangers. Offenses that were witnessed by other cadets were also more likely to be reported than those witnessed only by the respondent.

Several interactions were also significant. In terms of the two-way interactions, cadets had significantly stronger reporting intentions when the offenses were relatively severe and others had witnessed the event; much lower reporting intentions were noted for the low severity, low turn-in condition ($F(2, 728) = 7.90, p < .000$). The closeness $\times$ turn-in interaction term was also significant ($F(1, 364) = 15.95, p < .000$). Cadets reported stronger reporting intentions when a violation was committed by a relative

stranger and others had also witnessed the offense; lower reporting intentions were noted for the high closeness, low turn-in condition. Finally, all three within-subjects factors acted jointly to influence cadets' reporting intentions ($F(2, 728) = 98.75, p < .000$). The weakest intended action was in response to a low severity violation in which the observer and violator were emotionally close and there were no other witnesses to the policy infraction. Conversely, the strongest intended response was associated with a high severity violation in which the observer and violator were not emotionally close and others had also observed the policy infraction (see Figure 1).

Figure 1: Mean Severity of Response Ratings by the Three Situational Factors



To assess the effects of the demographic variables, a series of 1 between-, 3 within-subjects analyses were conducted. The strongest between subjects effect was noted for class year ($F(1, 361) = 11.30, p < .001$). Interestingly, this relationship was a curvilinear one; the mean responses for freshmen and seniors were higher than those for sophomores and juniors. A significant between-subjects effect for cadet squadron was also found ($F(3, 361) = 3.31, p < .02$), which implies that the social context in which cadets are immersed after entering the Cadet Wing influenced reporting intentions. There also appeared to be a linear effect for religiosity ($F(4, 360) = 2.56, p < .039$). Cadets reporting greater religiosity also reported they intended to take stronger actions across the scenarios. Finally, the reported MPA of cadets also was associated with differences in the mean intended action scores. This effect was also curvilinear; cadets with very low or very high MPAs reported stronger intended actions than did cadets with moderate MPAs.

Discussion

Intent to tolerate organizational policy or code violations, like most social acts, appears to be a complex process which is influenced by many factors. The current study indicates that both situational and demographic factors influence intent to report such infractions. Situational factors, such as emotional closeness to the violator, severity of the offense, and the likelihood of being turned in for failure to report, influenced reporting intentions both independently and interactively. Several demographic variables also affected reporting intention ratings.

Several comments about the feasibility of using low fidelity simulations to assess adherence to an organization's formal ethics policy seem warranted. First, we believe this study demonstrates the feasibility of using scenarios to assess the situational factors associated with adherence to a organizational policies. This methodology makes it possible to systematically vary the situational factors related to whistleblowing while minimizing the repetitiveness of the instrument. This was achieved by constructing scenarios which included people or violations considered to be psychologically equivalent by observers.

Second, we think this methodology could be expanded to examine the situational factors which influence a leader's behavioral intentions. For example, scenarios might incorporate individuals and performance

levels considered psychologically equivalent to systematically examine some of the tenets of attribution theory regarding interpersonal relationships, internal versus external attributions, and the decision to administer punishment (Mitchell & Wood, 1980). Perhaps this methodology could also be used to determine whether good and poor leaders attend to different situational factors.

Third, it may be possible to use the results of low fidelity simulations as a part of an organizational policy training program. This methodology avoids many of the ethical concerns associated with the laboratory and field studies of immoral or unethical behavior, and designing a training program to fit with the results of the low fidelity simulation may give people the skills to deal with difficult moral dilemmas at work. Although the utility of these programs has yet to be determined, the potential savings associated with the reduced incidence of employee unethical behaviors may be substantial.

In conclusion, low fidelity simulations may be a useful methodology for investigating a number of organizational phenomena without subjecting individuals to actual policy infractions. This study demonstrated that situational variables can affect an observer's toleration of another's unethical behavior. Future research should examine the generalizability of this methodology to the study of other organization phenomenon.

References

- Blasi, A. (1980). Bridging moral cognition and moral action: A critical review of the literature. Psychological Bulletin, 88, 1-45.
- Gordon, L. (1990, November 23). Rutgers students cheat fairly often. The Denver Post, p.8.
- Hegarty, W.H., & Sims, H.P. (1979). Organizational philosophy, policies, and objectives related to unethical decision behavior: A laboratory experiment. Journal of Applied Psychology, 64, 331-338.
- Kochanski, D. (1990). Contrails: The Air Force Academy Cadet Handbook (Vol. 36). United States Air Force Academy.
- Malinowski, C.I., & Smith, C.P. (1985). Moral reasoning and moral conduct: An investigation prompted by Kohlberg's theory. Journal of Personality and Social Psychology, 49, 1016-1027.
- Mitchell, T.R., & Wood, R.E. (1980). Supervisor's responses to subordinate poor performance: A test of an attribution model. Organizational Behavior and Human Performance, 25, 123-138.
- Motowidlo, S.J., Dunnette, M.D., & Carter, G.W. (1990). An alternative selection procedure: The low fidelity simulation. Journal of Applied Psychology, 75, 640-647.
- Sackett, P.R., Burris, L.R., & Callahan, C. (1989). Integrity testing for personnel selection: An update. Personnel Psychology, 37, 221-245.
- SPSS, Inc. (1988). SPSS-X User's Guide (3rd ed.). Chicago, IL: Author.
- Trevino, L.K. & Youngblood, S.A. (1990). Bad apples in bad barrels: A causal analysis of ethical decision-making behavior. Journal of Applied Psychology, 75, 378-385.

Authors' Note

The views expressed in this paper are those of the authors and do not necessarily represent the views of the United States Air Force Academy or any other governmental agency.



Perceived Job Security as a Function of Prior Military Service

Michael D. Matthews, Ph.D.
Drury College
Charles N. Weaver, Ph.D.
St. Mary's University

Abstract

Using national survey data, workers with prior military service were compared to workers who had never served in the military with respect to perceived job security. Results indicated no difference between the groups in expectation of losing their job within the next year. However, when asked how difficult it would be to locate a comparable job if forced to do so, respondents with prior military service indicated greater expected difficulty than did non prior service respondents. Implications of military service on civilian adjustment are discussed.

One of the benefits that is implied to accrue from military service is improved job skills, which should result in desirable outcomes for persons with prior military service as they enter the civilian labor force. Indeed, the connection between military experience and job skills forms the cornerstone for many recruitment campaigns by the armed services. Of course, other positive outcomes are also thought to follow military service. A recent issue of the Journal of Military Psychology (Gade, 1991) contains reviews the literature dealing with many possible consequences of military service, including post-service emotional well-being and the stability of subsequent family life.

However, with respect to work adjustment, the evidence concerning the effect of military service is mixed. In a major study, Laurence, Ramsberger, and Gribben (1989) compared workers with prior military service to those lacking such experience with respect to several occupational variables. Interestingly, little evidence of a positive effect of prior military service was found. In fact, veterans were found to have higher rates of unemployment, earn less money, and to have lower educational attainment than a matched group of non veterans. However, the generality of this study is questionable, because the prior military service sample consisted only of recruits with low aptitudes. Although this sample selection was consistent with the aims of the study, which was to look at the effects of military service among disadvantaged persons, it limits the degree to which the results may be applied to higher aptitude personnel. The latter may be more characteristic of the all volunteer force.

In a study using a nationally representative data base and including prior military personnel from all aptitude levels, Matthews and Weaver (1991) recently examined the relationship between prior military service and work attitudes. Specifically, their criterion variables were global job satisfaction and preferred job characteristics. Matthews and Weaver reported a significantly greater percentage of civilian employees with prior military service reported being "very satisfied" with their jobs than did comparable employees who had not served in the military. Moreover, workers with prior military service were more likely to express preference for intrinsic job characteristics, such as important, meaningful work than their non prior service counterparts. Similarly, veterans were less likely to report a preference for extrinsic variables, such as high pay. Thus, workers with prior military service in Matthews and Weaver's sample appear to enjoy their work more, and to be more highly motivated by intrinsic, rather than extrinsic, factors. This would represent a very desirable consequence of military service.

The purpose of the current study is to expand Matthews and Weaver's (1991) investigation to another work related criterion--job security. If military service provides a person with occupational advantages through vocational training, discipline, and character building, then a person with these advantages may feel more secure in their civilian occupation. To the extent this is true, it would lend further credibility to the argument that military service provides an enriching experience with positive vocational outcomes for those who possess it.

92-17666



Method

Data from the 1972 through the 1990 General Social Surveys served as the data base for this study. These door-to-door surveys (see Davis & Smith, 1990), conducted annually since 1972 except 1979 and 1981 by the National Opinion Research Center at the University of Chicago, are representative of the entire noninstitutionalized English-speaking population of the contiguous United States. All General Social Survey samples are of standard multistage cluster design, being modified probability for 1972 through 1974 and full probability for 1977 through 1990. For 1975 and 1976, the samples were half full and half modified probability, and comparison of data from the two halves reveals no important differences attributable to differences in sample design (Stephenson, 1979).

Male workers employed full-time who had served in the active duty military for two years or more (N=2260) were compared to persons with no active duty military experience (N=1464). The first criterion question asked: "Thinking of the next 12 months, how likely do you think it is that you will lose your job or be laid off--very likely, fairly likely, not likely, or not likely at all." The second criterion question asked: "About how easy would it be for you to find a job with another employer with approximately the same income and fringe benefits you now have? Would you say very easy, somewhat easy, or not easy at all?" Branch of service--Air Force, Navy, Army, National Guard, or Marines--was also identified.

The Chi Square test was used for making statistical inferences, and the standard rejection region of .05 was selected.

Results and Discussion

Table 1 summarizes responses to the first criterion question dealing with the perceived likelihood of losing one's job in the next 12 months. No significant differences are noted between the two groups with respect to this question, with close to two thirds of all respondents indicating little possibility of losing their job within the next year. Table 2 summarizes responses to this question as a function of branch of service. Again, no statistically significant relationship is noted. However, there is a trend for persons who had been in the Navy to perceive greater job security than persons who had served in any of the other branches.

Table 1

JOBLOSE by Veteran Status

Likelihood	NON PMS	PMS
Very Likely	6.2%	5.0%
Fairly Likely	5.3%	6.3%
Not Too Likely	23.1%	23.3%
Not At All Likely	65.3%	65.4%
	(N=2005)	(N=1392)

Table 2
JOBLOSE by Branch of Service

Branch					
	AF	NAVY	ARMY	NGUARD	MARINES
Very	5.7%	4.3%	5.7%	9.1%	6.8%
Fairly	5.7%	6.5%	8.0%	18.2%	3.4%
Not Too	25.5%	18.4%	22.0%	13.6%	23.7%
Not at All	63.6%	70.8%	64.3%	59.1%	66.1%
	(N=140)	(N=185)	(N=460)	(N=22)	(N=59)

Table 3 summarizes responses to the criterion question dealing with the ease of finding a comparable position. A significant difference in the responses of persons with and without prior military service is evident (Chi Square = 36.70; df = 2; $p < .01$). Respondents who had served in the military were substantially more likely to indicate it would not be easy to locate a comparable job than non prior service respondents. Moreover, as seen in Table 4, there are significant differences (Chi Square = 16.86; df = 8; $p < .05$) in the responses of persons who had served in different branches of the military to this question. In general, it appears that workers who had served in the Navy or the Marines perceive the least difficulty in securing a similar job, while those in the National Guard (however, note the small N) perceive the greatest difficulty. Former Army and Air Force personnel fall in an intermediate range, and indicate a similar perception with respect to the ease of finding a comparable job.

In interpreting these results, it appears that prior military service has little or no effect on job security in terms of expectation of losing one's job in the next year. Factors which contribute to an expectation of losing a job may have more to do with the general economic climate or occupational grouping than a person is employed within than do intrapersonal variables. However, an analysis of the relationship between prior military service and job security within clusters of civilian occupations might reveal some differences. For example, if military service produces a more skilled worker, then those having such experience who are employed in the trades subject to downturns during bad economic times might perceive greater security than those lacking military experience.

Table 3
JOBFIND by Veteran Status

JOBFIND	NON PMS	PMS
Very	28.0%	23.5%
Somewhat	31.6%	25.6%
Not Easy	40.4%	50.9%
	(N=1977)	(N=1385)

Table 4
JOBFIND by Branch of Service

JOBFIND	AF	NAVY	ARMY	NGUARD	MARINES
Very	21.6%	26.1%	19.9%	9.5%	39.0%
Somewhat	27.3%	23.9%	28.0%	19.0%	16.9%
Not Easy	51.1%	50.0%	52.1%	71.4%	44.1%
	(N=139)	(N=188)	(N=457)	(N=21)	(N=59)

The finding that workers with prior military service perceive a greater difficulty in locating a comparable job may be variously interpreted. One could argue that this is a negative outcome of military service in the sense that these respondents feel restricted in their future job opportunities. On the other hand, this could be interpreted as a positive outcome. It may be that prior military personnel have better jobs, in general, than those lacking military service. Consequently, they would be likely to perceive greater difficulty in locating a job with similar pay and benefits than someone employed in a marginal setting. The interpretation of these data is further complicated when differences noted among the services are examined. What is it about serving in the Navy or the Marines that makes respondents more likely to perceive that it would be relatively easy to obtain a similar job? Do they have better, or more general training and experience while in the service than do persons in the other branches?

In order to answer some of the questions raised here, a further analysis of these data examining what types of civilian occupations are held by the prior military service sample versus the non prior service sample would be fruitful. For instance, are persons with prior service more likely to be employed in certain types of occupations and, if so, are there still differences noted between those with and without prior service? Other control variables such as age, education level, and so forth should also be considered before general conclusions concerning the role of prior military service on civilian work adjustment are drawn.

In summary, the current data along with those reported earlier by Matthews and Weaver (1991) suggest that differences in work attitudes do exist between workers with prior military service and those without military service. Full interpretation of these differences await further analyses in which additional control variables are specified.

References

- Davis, J.A., & Smith, T.W. (1990). General Social Surveys, 1972-1990: Cumulative Codebook. Chicago: National Opinion Research Center.
- Gade, P. A. (1991, ed.) Special issue: Military service and the life course perspective. Journal of Military Psychology, all of Volume 3.
- Laurence, J.H., Ramsberger, P.F., & Gribben, M.A. (1989). Effects of military experience on the post-service lives of low aptitude recruits: Project 100,000 and the ASVAB misnorming. (Final Report #89-29). Alexandria, VA: Human Research Organization.
- Matthews, M.D., & Weaver, C.N. (1991). Work attitudes of employees with prior military service. Proceedings of the 33rd Annual Meeting of the Military Testing Association. San Antonio: Tx: Armstrong Laboratory Human Resources Directorate & USAF Occupational Measurement Squadron, in press.
- Stephenson, C.B. (1979). Probability sampling with quotas: An experiment. Public Opinion Quarterly, 43, 477-496.



Age Group Differences and Estimated Frequencies of the MBTI Types:
Proposed Changes (M I N INTJ Forever?)

William H. Cummings
Defense Systems Management College
Ft. Belvoir, Virginia 22060

Abstract

An analysis of key tables in the Atlas of type tables for the Myers-Briggs Type Indicator showed strong, unexpected age group effects, particularly for the JP dimension. These effects are not predicted by--and may contradict certain aspects of--traditional type theory. These findings prompted a consequent recalculation of MBTI norms. Proportions of extraverts and sensors were found to be considerably lower than previously reported, while the proportion of judgers was somewhat higher.

The Myers-Briggs Type Indicator (MBTI) is a personality assessment instrument that enjoys solid validity and reliability, as well tremendous popularity in recent years. In fact, DeVito (1985) estimates that "the MBTI is probably the most widely used instrument for non-psychiatric populations in the areas of clinical, counseling, and personality testing" (p. 1030). However, for all its durability and popularity, the emphasis on the MBTI remains focused on practical applications rather than research. Interest and application have clearly surpassed what we know about this valuable instrument. This paper addresses two important--and related--research questions:

- Are peoples' MBTI types invariant over time, as type theory implies?
- Do the widely quoted MBTI norms reflect the best available data? In other words, are the MBTI norms accurate, to the best of our knowledge--and particularly in light of the previous question?

Designed as an operationalization of Carl Jung's theory of psychological types, the MBTI describes individuals in terms of four dichotomous indices. These four indices reflect individuals' basic preferences which, in turn, direct their key behaviors. The indices are:

Extraversion-Introversion (EI): Does the individual focus his/her perception and judgments primarily on actions and persons in the outer world (extraversion, "E"), or on concepts in the inner world (introversion, "I")?

Sensing-Intuition (SN). Does the individual perceive information primarily in terms of concrete facts and details (sensation, "S"), or in terms of meanings, concepts, and relationships (intuition, "N")?

Thinking-Feeling (TF). Does the individual make judgments and decisions primarily on the basis of logic and objective analysis (thinking, "T"), or on the basis of personal, social, and subjective values (feeling, "F")?

Judgment-Perception (JP). In dealing with the outer world, does the individual prefer order, closure, and structure (judgment, "J"), or flexibility, openness, and a free flow of information (perception, "P")?

True to Jungian theory, the MBTI "types" the individual on each index, and the four indices are combined to yield sixteen composite types. Thus, each individual can be described as an ESTP, INTJ, etc. Also true to Jungian theory, it is widely assumed (though seldom tested) that these preferences are fixed and should not vary with age (e.g., Kroeger & Thuesen, 1988; Myers & McCaulley, 1985). Thus, it is common practice to say an individual "is" an ENFJ, rather than "behaving as an ENFJ," or "high in ENFJ-ness."

92-17667



A second issue of widespread interest to MBTI practitioners is the normative data on the MBTI, i.e., the approximate frequency of each type in the U.S. population at large. A number of sources estimate these percentages:

- Approximately 70-75% of the population prefer extraversion, with the remaining 25-30% preferring introversion.
- Approximately 70-75% prefer sensation to intuition.
- Approximately 50% prefer thinking to feeling. However, the TF scale also shows a reliable sex difference, with 60-70% of males preferring thinking, and 60-65% of females preferring feeling.
- Approximately 50-65% of the population prefer judgment to perception (Jeffries, 1991; Keirsey & Bates, 1984; Kroeger & Thuesen, 1988; McCaulley et al., 1985; Myers, 1962, cited in Myers & McCaulley, 1985; Myers & McCaulley, 1985).

Since 1971, the Center for the Applications of Psychological Type (CAPT) has received 232,557 records for entry into the MBTI data base. This data base--classified in the Atlas of type tables (Macdaid et al., 1986) by age group, sex, education level, occupation, and data source--provides a rich research resource which, in turn, will help us address the related questions of longitudinal stability of MBTI type and accuracy of the widely cited norms.

Method

This study focused on the sixteen tables from the Atlas of type tables that are classified by age and sex of the subject. The tables are broken out into eight age groups for each sex: age 15-17, 18-20, 21-24, 25-29, 30-39, 40-49, 50-59, and 60+. The tables reflect the CAPT-MBTI Data Bank of MBTI records that were submitted to CAPT for scoring between 1971 and December 1982 and contain data for 36,713 males and 46,443 females. (An additional 5% of the records contained no age data and thus are not represented in the tables.)

The following tables were used for supplementary analyses: adult high school dropouts, adult college graduates, traditional age high school students, traditional age junior high school students, and the CAPT Data Bank total population. Also used were data obtained by CAPT from the SRI International study on American Values and Life Styles. The SRI data were of particular interest, since they were collected as part of a mail-out survey, rather than in the traditional face-to-face administration.

Results and Discussion

Table 1 gives the percentages of extraverts (E), sensors (S), thinkers (T), and judgers (J) for both males and females for all age groups. (The percentages of introverts, intuitors, feelers, and perceivers are not given here, since they are simply ipsative reflections of the four corresponding preferences.) Also shown are the chi-square value and the corresponding phi coefficient for the variation between the different age groups within each sex group. (The phi coefficient is particularly helpful in this instance in interpreting the magnitude of the age effects, since the chi-square statistic may be biased by large Ns, approaching "perfect" power.)

Table 2 gives the percentages of E's, S's, T's, and J's for males versus females. Values in Table 2 are based on the age-group percentages in Table 1, adjusted for the proportional size of each age group in the U. S. population, as shown in the Statistical abstract of the United States: 1990. For example, the Female Age 60+ group contained a relatively small proportion of records but received a 23.80% weighting, reflecting the true proportion of this age

group among females in the U. S. population. This weighting technique should yield different--and more accurate--estimates of the population percentages of the MBTI preferences than previously given. This will be particularly true if there are age group differences in MBTI preferences.

The data in Tables 1 and 2 reveal some interesting findings, both confirming and refuting various long-held assumptions:

1. Table 2 further confirms the established finding of no/minimal sex differences, with the exception of the well-documented TF difference. (Since this difference is among the strongest and best established MBTI effects, the phi value of .298 serves as a benchmark for the other results examined here.)

2. Table 1 casts considerable doubt on the Jungian/type theory assumption of typological invariance across age groups. In fact, all four dimensions show a significant age group effect for both males and females (though this effect is small in some specific instances). In particular, the JP dimension shows a strong, almost linear increase in percent of J's for both males and females. This change approximates thirty percentage points, from the youngest to the oldest groups. This effect alone would be enough to cast doubt on any assumption of typological invariance over time. However, the other three dimensions also reveal age group effects:

a. Females show a distinct decrease in percent of E's, some fifteen percentage points from youngest to oldest age group.

b. Both males and females show a curvilinear effect over age groups in percent of S's, with the most S's in the younger and older age groups.

c. Both males and females show a curvilinear effect in percent of T's, with the most T's in the middle age groups.

At least four alternative explanations readily account for these age group differences: (a) a true developmental change (e.g., people become more J-oriented over time), (b) a differential willingness to self-report certain behaviors over time, (c) generational-based, value-programming effects (e.g., members of older age groups may have been value-programmed with more traditional, J-type values), and (d) sampling differences between younger and older age groups represented in the CAPT Data Bank.

The data here do not help us determine which explanation is correct. However, the data certainly cast clear doubt on a simple assumption of

TABLE 1
CAPT Data Bank Percentages of MBTI Preferences by Age and Sex

Age:		15-17	18-20	21-24	25-29	30-39	40-49	50-59	60+	$\chi^2(df=7)$	Phi
%E:	M	51.6	52.2	51.7	46.5	48.7	49.1	50.9	47.5	54.54 ^a	.038
	F	61.6	60.8	56.5	50.1	49.4	51.5	50.4	47.2	506.79 ^a	.104
%S:	M	56.2	58.9	53.1	48.7	51.7	56.6	60.2	64.4	206.81 ^a	.075
	F	57.4	62.2	56.6	52.3	52.1	51.9	55.9	61.1	337.60 ^a	.085
%T:	M	62.0	62.3	63.3	67.0	71.2	73.5	71.4	67.7	306.61 ^a	.091
	F	32.3	31.0	37.2	43.3	42.8	39.9	37.7	36.3	492.01 ^a	.103
%J:	M	45.7	51.8	59.3	61.4	67.9	71.4	72.9	74.8	1178.01 ^a	.179
	F	48.1	56.9	61.1	62.2	64.3	66.0	67.7	75.4	646.39 ^a	.118

^ap < .05

TABLE 2
CAPT Data Bank Percentages of MBTI Preferences by Sex

	Sex: Male	Female	$\chi^2(df=1)$	Phi
%E:	49.2	51.2	34.25 ^a	.020
%S:	56.3	55.9	1.50	.004
%T:	68.7	38.7	7386.87 ^a	.298
%J:	66.5	65.9	3.62	.007

^ap < .05

typological invariance with age. They also call for two necessary actions:

- a. Longitudinal and related research to investigate the possibility of systematic developmental changes in type, and
- b. Suspending any judgment of typological invariance. MBTI users should inform their clients the available data do not support an assumption of typological invariance and may indicate otherwise.

3. Regardless of cause, the age effect on virtually every dimension signals the need to re-estimate the MBTI norms. This need becomes imperative, given that early norms were based largely on high school samples, based in turn on an assumption of psychological invariance. Furthermore, no norms to date have adjusted the obtained data for the proportional size of each age group in the U. S. population (as done in Table 2).

Table 3 presents data from the current study--as well as various other data bases--in an attempt to recalibrate the MBTI norms. Each other data base included in Table 3 represents an extreme group in terms of one or more MBTI dimensions. For example, the Adult High School Dropout group (Column 4) represents a group that, according to MBTI theory, should be high in S's and P's. These samples allow us not only to estimate the "best" population percentage for each dimension, but to establish upper and lower bounds as a logic check. Therefore, the highest and lowest value for each row in Table 3 are marked as "U" (Upper Bound) and "L" (Lower Bound), respectively. Any norms--from whatever data source--that exceed or even approach these bounds should be viewed with extreme skepticism.

The preference percentages from the current analyses fall well within these upper and lower bounds. However, at least one source of bias is likely to remain in these data: Standard MBTI data collection yields an undetermined bias toward college-educated professionals working in organizational settings.

TABLE 3
Preference Percentages for Selected MBTI Data Bases

	Curr. Study	CAPT Data	SRI Study	Adult Drop.	Coll. Grads	15-20 YrOlds	50+ YrOlds
%E:	50.2	53.2	40.4(L)	46.4	49.1	56.5(U)	48.5
%S:	56.1	54.1	75.9(U)	70.7	48.2(L)	58.7	61.1
%T (Male):	68.7	n/a	74.7(U)	69.7	70.9	62.2(L)	69.1
(Fem.):	38.7	n/a	34.0	42.0	43.0(U)	31.6(L)	36.8
(M-F):	53.1	42.3(L)	50.4	55.8	56.9(U)	46.9	52.9
%J:	66.2	57.0	66.2	67.7	68.2	50.7(L)	73.5(U)

Therefore, these data are probably somewhat biased in favor of I-, N-, and J-types. With this precaution in mind, the following population percentages are recommended for use with the MBTI:

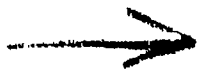
% Extraverts:	50%	/	% Introverts:	50%
% Sensors:	55-60%	/	% Intuitors:	40-45%
% Thinkers (Male):	65-70%	/	% Feelers (Male):	30-35%
% Thinkers (Female):	35-40%	/	% Feelers (Female):	60-65%
% Thinkers (M+F):	50%	/	% Feelers (M+F):	50%
% Judgers	65%	/	% Perceivers:	35%

These recommended percentages approximate previously given percentages for some dimensions (e.g., TF) but differ widely for other dimensions (EI, SN). The differences may be due to the strong reliance of some earlier studies on high school samples, with a corresponding assumption of typological invariance. Furthermore, the early work of Myers (1962; cited in Myers & McCaulley, 1985) remains influential; Myers estimated large percentages of extraverts and sensors and a relatively small percentage of judgers.

In conclusion, this analysis of data from the Atlas of type tables classified by age and sex showed unexpected differences between age groups. Although the current data do not help us determine the exact cause for these effects, the findings do cast clear doubt on a simple assumption of typological invariance. Furthermore, these findings prompted a recalculation of MBTI norms. These new norms indicate a smaller percentage of extraverts and sensors and a somewhat greater percentage of judgers than did most previous normative studies. It is recommended that MBTI users adopt these new norms, while withholding any assumption of typological invariance with age.

References

- Bureau of the Census. (1990). Statistical abstract of the United States: 1990 (110th ed.). Washington, D. C.: U. S. Government Printing Office.
- DeVito, A. J. (1985). Review of the Myers-Briggs Type Indicator. In J. V. Mitchell, Jr. (Ed.), The ninth mental measurements yearbook (pp. 1030-1032). Lincoln, NE: Buros Institute of Mental Measurements.
- Jeffries, W. C. (1991). True to type. Norfolk, VA: Hampton Roads.
- Keirsey, D., & Bates, M. (1984). Please understand me. Del Mar, CA: Gnosology Books.
- Kroeger, O., & Thuesen, J. M. (1988). Type talk. New York: Dell.
- Macdaid, G. P., McCaulley, M. H., & Kainz, R. I. (1986). Atlas of type tables. Gainesville, FL: Center for Applications of Psychological Type.
- McCaulley, M. H., Macdaid, G. P., & Kainz, R. I. (1985). Estimated frequencies of the MBTI types. Journal of Psychological Type, 2, 3-9.
- Myers, I. B., & McCaulley, M. H. (1985). Manual: A guide to the development and use of the Myers-Briggs Type Indicator. Palo Alto: Consulting Psychologists Press.



AD-P006 935



The Myers-Briggs Type Indicator and
Acquisition Management

Lt Col Carl Bryant Ph.D.
Defense Systems Management College

92-17668



Abstract

This paper describes the Myers-Briggs typology of over 2,800 students who attended the Defense Systems Management College (DSMC) since 1983. It further describes the potential contributions and pitfalls that each type brings to an organization.

The Defense Systems Management College (DSMC) was created to provide acquisition education, research, information dissemination, and oversight of the total Department of Defense acquisition education program. To meet this goal the Department of Managerial Development has since 1983 administered the Myers-Briggs Type Indicator (MBTI) to students who attended the Program Management Course (PMC). The MBTI was administered to provide these students with a way of understanding individual differences and utilizing this knowledge in their Total Quality Management efforts. This paper describes the psychological type of a representative sample of more than 2,800 students who work in the acquisition career field and who attended the PMC since 1982.

Method

Subjects

Participants in this study consisted of 2,818 military officers, civil servants and defense contractors who attended the PMC between 1982 and 1991. This sample consisted of 804 U.S. Army personnel, 786 U.S. Navy personnel, 854 U.S. Air Force personnel, 123 U.S. Marines, 169 from defense-related industry and 12 from the U.S. Coast Guard. Due to the small number of Coast Guard students, crosstabulations for that group were not preformed.

Procedure

All students were given the Myers-Briggs Type Indicator Form G on the first or second day of the PMC. According to Briggs-Myers and McCaulley (1989), p.3), the main purpose of the MBTI is to identify four basic preferences. The four preferences are:

Extroversion-Introversion (EI). The EI index is designed to reflect whether a person is extroverted or introverted. Extroverts are oriented primarily toward the outer world; thus they tend to focus their perception and judgment

on people and objects. Introverts are oriented primarily toward the inner world and thus tend to focus their perceptions and judgments upon concepts and ideas.

Sensing-Intuition (SN). The SN index is designed to reflect a person's preference between two opposites ways of perceiving; one may rely primarily upon the process of sensing, which reports observable facts or happenings; or one may rely upon the less obvious process of intuition which reports meaning, relationships, and/or possibilities that have been worked out beyond the reach of the conscious mind.

Thinking-Feeling (TF). The TF index is designed to reflect a person's preference between two contrasting ways of judgment. A person may rely primarily on thinking to decide impersonally on the basis of logical consequences, or a person may rely primarily on feeling to decide primarily on the basis of personal or social values.

Judgement-Perception (JP). The JP index is designed to describe the process a person uses primarily in dealing with the outer world, that is, with the extroverted part of life. A person who prefers judgment has reported a preference for using a judgment process (either thinking or feeling) for dealing with the outer world. A person who prefers perception has reported a preference for using a perceptive process (either sensing or intuition) for dealing with the outer world.

Results

The participants in this study were found to be very similar to the people McCaulley (1990) identified as managers. However, the ENTJs which she describes as top executives comprised only 9.4% of this sample.

What implications might this have for TQM efforts in the acquisition career field? To begin, 56.1% of the students in this study were STs. They want clear procedures, which seldom change and that everyone follows (often without question).

What are the implications of having STs dominate an organization such as a Program Management Office, which has been characterized as being in an almost constant state of change? How do you energize this portion of the work force and its leadership that perhaps is operating in the environment they find least rewarding and perhaps most stressful? Point two of Deming's fourteen points ask organizations to "adopt the new philosophy"

(Walton, 1986). This is an area in which ST have considerable trouble. Adoption of TQM within the acquisition career field may be seriously hampered by the STs resistance to change. This question is worthy of further investigation.

The SFs comprise only 5.5% of this sample. Who takes care of their need for inclusion in the Program Management Office? It is possible that in an environment dominated by STs they perceive themselves as different, not fitting in, and unappreciated. If so, achieving Deming's first point, "creating constancy of purpose for the improvement of product and service" may be very difficult.

The NFs make up 6.3% of this sample. How is the purpose of the Program Management Office communicated? Is any attention paid to the service of the country or humankind? If not, how does a Program Manager get this group to "buy into" his or her goal? Walton (1986) states that Deming's seventh point is "institute leadership" (p.70). She further states "It is the responsibility of management to discover the barriers that prevent workers from taking pride in what they do" (p.70). Paying attention to the feelings and aspirations of NFs will be critical to any manager who hopes to get them fully on his or her team.

Finally, the NTs comprise 32% of this sample. For them the vision of the organization and where it is going in the long run is very important. Is the plan clear? If not, this group will detect inconsistencies, and their energy will possibly be diverted to other areas of interest. The NTs appear to be an excellent group to utilize as the organization attempts to implement Deming's thirteenth point Institute a vigorous program of education and retraining. They are energized by possibilities and by learning. This could be an important assess for any manager working in the acquisition environment.

One other which must be addressed is related to diversity and to predicted changes in the United States workforce. If the demographic changes which are predicted by Johnston and Packer (1987) occur, far greater emphasis will have to be placed on team building in the Program Management Office of the future. These changes include more women entering the workforce and minorities comprising a larger share of new workers. The sample described in this study is more than 92% male. Of the men described in this study, approximately 95% are white. How successful this largely homogenous population will be at adapting to a changing work force is a matter of no small debate. What is clear, however, is that change will be occurring. The question of the resistance to change by this population is evidenced by the statements of many PMC students. Recent PMC students have reported a great deal of resistance to the implementation of Total Quality Management in their organization. If this resistance is, in fact a characteristics of the acquisition environment, creating successful work groups may become more and

more difficult. The time to address these issues is now. Hopefully, this paper will be of assistance in that effort.

References:

- Briggs-Myers, I., & McCaulley, M. H. (Eds.). (1989). A guide to the development and use of the Myers-Briggs Type Indicator (fifth printing). Palo Alto, CA: Consulting Psychologist Press.
- Johnston, W. B., & Packer, A. E. (1987). Work and workers in the year 2000. In Hudson Institute, Workforce 2000 (pp. 75-104). Indianapolis, IN: Hudson Institute.
- McCaulley, M. H. (1990). The Myers-Briggs Type Indicator and Leadership. In K. E. Clark & M. B. Clark (Eds.), Measures of leadership (pp. 381-418). Greensboro, NC: Leadership Library of America
- Walton, M. (1986). The Deming Management Method. New York: Putnam.



AD-P006 936



Visual Representation Study

92-17669



CLC Blair W. Conroy
Captain Jay A. Horn, M.S.
Major Paul D. Grunzke, Ph.D.
USAF Academy
Christopher D. Wickens, Ph.D.
University of Illinois at Champaign-Urbana

Abstract

The purpose of this study is to define the most effective data presentation style for a given type of task. Three display modes are proposed in this study: Color, 3-Dimensional bars and lines or 2-Dimensional. However, only the color condition has been tested. The study requires subjects to answer questions concerning the relative levels of each graphic dimension. However, one type of question requires integrative thinking (trend analysis) and the other requires focused thinking (point reading). Selection of the different display modes are based on the principle of proximity compatibility or providing an integrated format for integration based questions. The results support our hypotheses about the effectiveness of graphic displays. We constructed our own fictional database of information regarding cadet physical performance and designed balanced questions to test two types of information processing. Our test only covered the medium integration display format--color, however, this condition alone yielded mean reaction times that were greater for trend questions and lower for focused questions. This supports the proximity compatibility principle which is the foundation of our hypotheses. We also found that reaction times increased across our four levels of graph complexity: 2x2, 2x4, 4x2, 4x6.

Visual representation is used in every conceivable medium in every aspect of our lives. In several cases, the method of representation is selected without any statistical support. The current study examines visual representations or graphical displays that summarize data.

The main school of thought that inspired our hypothesis regarding visual representation effectiveness is the theory of the Proximity Compatibility Principle (PCP) described by Goettl, Wickens and Kramer (1991). This theory presupposes that different types of visual displays are best for different types of tasks.

The PCP uses both data describing how people encode information and visual searching techniques. The PCP also suggests guidelines for different task related displays.

According to Merwin and Wickens (1991), tasks that require mental integration should use integrated formats (3-D graphs) while tasks that require focused attention should use separate formats (2-D panel displays).

The current study seeks to provide empirical evidence for graphic display effectiveness. The main hypothesis is that questions requiring mental integration are best answered using an integrated display format while questions requiring focused comparisons are best answered using a separated display format. As shown in figure 2, three display formats for a three variable data base are included in the full experimental design, a 3-D perspective display (most integrated), a set of 2-D planer displays (least integrated), and a planar display using color to represent the third dimension. At the time of this writing, only data on the third condition has been collected.

Method

The current study was conducted at the United States Air Force Academy. Four cadets were randomly selected to participate in a study of computer generated visual representation effectiveness. Our report is the initial stage of the visual representation study.

Subjects

The four male subjects ranged in age from 20 to 23. Each subject had prior experience with both the computer used for testing and the sample data base concerning physical fitness test scores.

Figure 1

Color Graph

Class Year	Upper	Blue (500)	Gray (300)
	4 ^o	Red (100)	Blue (500)
		88	89
Graduation Year			

Apparatus

The colored graphs used were based on a hypothetical data base of physical fitness scores of cadets over a period of six years (See figure 1).

Table 1

Experimental Design

Integration	Complexity ¹	Questions
3-D	2x2 2x4 4x2 4x6	1 Focus/1 Trend
Color	2x2 2x4 4x2 4x6	1 Focus/1 Trend
Lines/2-D	2x2 2x4 4x2 4x6	1 Focus/1 Trend

Note. ¹ Refers to number of levels of PFT score x Class x Graduation Year. All conditions include five levels of PFT score.

All the graphs used were randomly generated and used fictional data concerning physical fitness scores. The variables represented in the displays were fitness score, graduation year, and class year (freshman through senior). Five colors were used to represent the five different levels of fitness scores: red 0-100, yellow 100-200, gray 200-300, green 300-400, blue 400-500. We chose these colors based on a pilot study of color preference in the Cadet Wing. The strongest color bias was tied to the order (highest to lowest) : blue, green, yellow and red. We added gray to provide a neutral fifth level. The dependent variables used were latency and accuracy. The independent variables used were complexity of the graphs, and focused vs. trend questions. Complexity was varied by the number of class levels represented in the data base and the number of years presented. Thus figure 1 shows an example of a "2x2" display presenting two class levels and two graduation years. Higher levels of complexity were created by expanding the number of classes (rows) to four, and the number of years (columns) to six. Table 1 presents the full range of experimental conditions. Only data in the first row are reported in this paper. The design ensured that each subject would get all four levels of complexity and the latency difference between complexity levels would be measured.

Procedure

The four subjects were allowed to practice for five minutes or until they felt comfortable with the color scheme. In addition, they were allowed to use a color chart showing the relative ranking of the five colors used. Next they each answered questions, respectively 55, 48, 45 and 45 for 10 minutes which amounted to 193 total trials recorded. The questions were randomly assigned to the subjects and the variations of questions worded positively and negatively were balanced for each subject.

Questions were modeled after three types used by Merwin and Wickens (1991): focused on one dimension of graduation year, focused on one year but integrated across classes, and focused on one year and integrated across scores. Some example questions are shown in table 2.

Table 2

Question Types

Integration Level	Difficulty	Example Question
Trend	Low	Do PFT scores tend to converge?
Trend	High	Do upperclass scores decrease while 4° scores remain constant?
Focused	Low	Did the 4°s in 88 score higher than upperclassmen in 88?
Focused	High	Did the 4°s score the lowest both years?

Results

Table 3 presents the reaction times recorded. As is evident from table 3, no difference was found between the levels of integration (focused vs. trend). The mean reaction times are 2.93 sec. and 3.15 sec. respectively. The analysis of variance confirmed that no significant difference existed between levels of integration ($F=.8758$ $p<.351$). Table three also reveals that reaction time increases with increasing complexity. This result is shown in both focused ($F=4.63$ $p<.004$) and trend ($F=2.94$ $p<.038$) questions.

Table 3

Mean Reaction Times

Integr. Level	Focus	2.93	2.27	3.24	3.19	3.42
	μ/s	1.42	.9	1.45	1.51	1.67
	Trend	3.15	2.62	2.94	3.29	4.21
	μ/s	1.86	1.97	1.68	1.54	1.88
		Total	2x2	2x4	4x2	4x6
Complexity						

Note. All times are in seconds. Complexity refers to class x graduation year.

Discussion

Current data are of course inconclusive with regard to the PCP because the PCP predicts a display x task (question type) interaction and, at present, only one display type was examined. The main hypothesis of our program of research is that performance on tasks requiring point selection (focused questions) will decrease as display integration increases while performance on tasks requiring information integration (trend questions) will increase with display integration. In the present experiment we only examined the mid-level of integration (color). This is where the PCP predicted few differences between the task levels. This interaction is shown in figure two along with the other predicted levels of performance. Our results are consistent with this view. The next step is to collect data on the remaining two conditions shown in figure 2. The importance of testing these two conditions is paramount since our paradigm demonstrated such strong complexity effects. The true empirical test of the PCP is yet to come.

References

- Carswell, M.C. (1991). Reading graphs: Interactions of processing requirements and stimulus structure. Percepts, concepts, and categories: The representation and processing of information. (in press).
- Goettl, B. P., Wickens, C. D. and Kramer, A. F. (1991). Integrated displays and the perception of graphical data. Ergonomics, 34, 1047-1063.
- Merwin, D. H. and Wickens, C.D. (1991). 2-D vs 3-D display for multidimensional data visualization: The relationship between task integrality and display proximity. (In press).
- Wickens, C. D. (1992). Engineering psychology and human performance. New York, NY: Harper Collins.



Distance in Multifunction Displays

Karen Seidler and Christopher D. Wickens
 Aviation Research Laboratory, Institute of Aviation
 University of Illinois, Savoy, IL

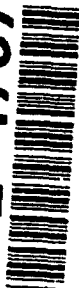
ABSTRACT

Three forms of "distance" metrics are defined between entries in a multifunction display: navigation distance (defined by key presses), organizational distance (defined by the hierarchical structure), and subjective distance, defined by the pilot's "mental model" of items similarity. These three distance metrics are contrasted in an experiment in which pilots must travel between 290 screens of an MFD containing flight information. Travel time was greater with longer navigation distance and longer subjective distance, but was actually shorter between screens that were more organizationally distant (i.e., in different branches of the structure). This effect is attributed to the role of pilot strategies in navigation.

The introduction of the computer into the flight deck has led to several uses. One of the most critical of these has involved using computers to support information access and availability. Two noteworthy examples here are the Electronic Library System (ELS, Curran, 1991), and the Multifunction Display (MFD, Reising & Curry, 1987). Both of these tools serve to store and make available electronically, that which would otherwise be accessed by document page turning (the ELS), or visual scanning across a wide display surface (the MFD). In both systems the designer's goal is to make information rapidly available. The question is, how to do so? In the following treatment, we adopt a spatial metaphor, to describe these electronic data bases emphasizing the concept of "distance" between different entries in the data base. The designer's task then must consider as one criterion, minimizing this distance.

We consider navigation distance to be the number of key strokes necessary to proceed from one screen (i.e., a "start") to another (i.e., a destination or "target"). One solution to minimizing navigation distance is simply to provide "direct access" commands to all entries. Hence the distance of any screen from any other is only a single keyboard entry. But as the number of screens grow large (several hundred in envisioned ELSs), the demands of remembering (or looking up) the direct access commands become excessive, particularly for the novice user. The alternative to direct access is to impose some hierarchical organizational structure on the data base, allowing the pilot to proceed down a menu structure to reach a desired screen (Norman, 1991). Here navigational distance would typically be defined by the number of "nodes" in the menu between a start and target screen. However, the menu system could provide navigational "shortcuts" e.g., by direct access keys, or by an option to immediately jump to the top level of the menu. In this sense we have now defined a second form of "distance", organizational distance which is defined in terms of the map of the formal structure of the hierarchy. For example, in Figure 1, the organizational distance from the COS runway characteristic to the entry describing the aircraft's flight characteristic is

92-17670



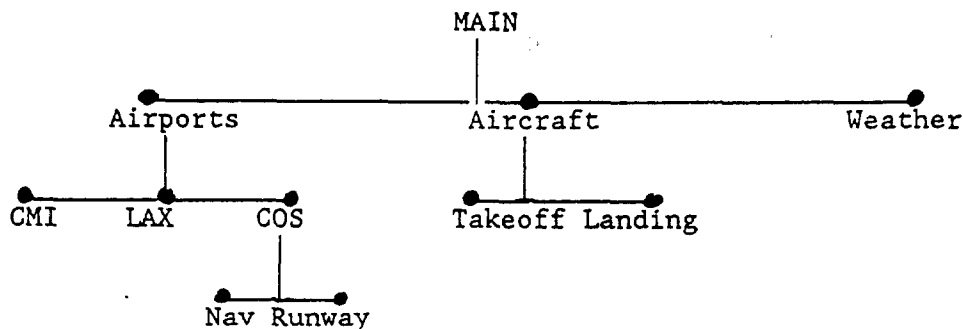


Figure 1

4. The navigational distance would be 1 if the pilot has direct access, or 2 if a "main menu" page is available.

There is yet a third distance metric in menu design, which is that based upon the pilot's "mental model" of what information screens are subjectively "close" together (Roske-Hofstrand & Paap, 1986). This subjective distance may vary from pilot to pilot or occasion to occasion. For example, a pilot preparing to land at COS will find the COS runway characteristics and aircraft landing parameters "close" to each other. One who is learning about aircraft performance characteristics will find landing and take off characteristics close. It is important to note that organizational distance imposed by the system designer, and subjective distance experienced by the user may be very different from each other. Under these circumstances, research has found that there is a substantial cost to performance caused by this disagreement (Roske-Hofstrand & Paap, 1986).

The goal of the current study is to contrast these three measures of distance, navigational, organizational, and subjective, in the pilot's use of a simulated MFD. Our interest was in the speed and mental workload with which pilot's could navigate from a "start" screen to a "target" screen when different navigational tools were provided and different degrees of agreement between organizational and subjective distance were compared.

METHOD

Fifteen instrument rated pilots were paid \$5.00/hour for their participation in the study. The task, programmed on SuperCard on a Macintosh Si computer, required pilots to navigate through a large (290 screen) hierarchical data base of flight information. The menu's top four categories defined information regarding aircraft systems, navigational planning, weather information, and aircraft performance. In the "go to" navigation task, pilots, viewing a given screen, would receive instructions to "go to" another screen, designated by its title. They also received a 5 digit number which was to be rehearsed and retained until the target screen was reached. Performance on this memory task was used to assess the mental workload of the traversal. Upon reaching the target screen, the 5 digit number was entered

and a new target screen was identified. In different conditions, subjects had two different navigational tools available. In the previous condition, subjects could only step up or down the menu structure one node at a time. That is, to move up the hierarchy subjects had to depress a "previous" screen key. In the "main" condition, subjects could type a "main menu" key which would immediately jump them up to the top level of the menu. By comparing navigation across the same organizational routes between these two conditions, the effects of navigation distance can be assessed. For a path between two nodes on different main menu categories, navigational distance with the main menu key will be shorter. By comparing two paths between the two conditions that have the same navigation distance, but different organization distance (i.e., within vs. between a main menu category), the effect of organizational distance can be determined.

Finally, prior to the experiment, a questionnaire was administered regarding specific trios of screens, to determine the relative subjective similarity between the three. The structure of the three screens in each trio was such that two (A and B) were within one main menu category, and the third (C), was within a different category but was suspected to be highly related subjectively to A or B. The purpose of the questionnaire was to assess this degree of relatedness. If the subjective relatedness was greatest between A and B, then subjective distance was assumed to be congruent with organizational distance (since A and B both lay within the same higher level node and were thus, by definition, organizationally "closer"). If subjective distance was closer between C and A or C and B, than between A and B, then the relation between the two distance metrics was said to be incongruent.

Following practice, all subjects encountered the same 114 screens across two 1 1/2 hours sessions. The navigational paths however were presented in different orders for different subjects.

RESULTS

The following presents the highlights of the results. More detailed presentation of the methods and results may be found in Seidler, Wickens, and Davis (1991). Across scenarios, a main effect of navigation distance was found. Not surprisingly, more nodes visited between start and target screen, led to longer retrieval time ($F = 33.47$; $p < .01$) and greater workload as revealed by the loss of memory accuracy.

The differential effect of navigation and organization distance was examined by comparing two sequences between the main and previous menu conditions, that imposed the same navigational distance, but different organizational distances. Sequences with greater organizational distances (using the main menu key) provided significantly shorter traversal times than those with shorter organizational distances. This unexpected finding was interpreted in terms of subject strategy. When starting at a low level screen it was found to be easier to simply jump to the top level of a menu and then work down (the main menu key), than to work up one step at a time, thinking after each step whether a high enough level has been obtained to work back down a different branch to the target screen. Correspondingly, within the trials using only the previous key, subjects were faster moving between screens in different categories than within the same category, a finding explained by the same logic.

The role of subjective distance was evaluated by comparing navigational time between those sequences in which subjective and organizational distances was congruent and incongruent. The incongruent sequences were found to take significantly more time to traverse, replicating the findings of Roske-Hofstrand and Paap (1986) findings regarding the importance of a user's mental model of the data base.

DISCUSSION

The present results document the importance of all three distance metrics. It is not surprising that navigational distance plays a role in search time. Each node visited requires a physical action (key press) and a decision as to which node to proceed to next. Decision time for the latter component will of course vary as a function of the menu "breadth" and the organization of options on the screen which will influence visual search time, an issue not considered here (but see Snowberry, Parkinson, & Sisson, 1983; Seidler, Wickens, & Davis, 1991).

The role of organizational distance was somewhat different from what had been anticipated. Longer organizational distances (i.e., between different main menu categories) led to shorter traversal times, when navigational distance was controlled for. This finding implicated the important role of strategy, and the greater difficulty that subjects encounter when moving up the menu toward an uncertain reversal point to initiate the movement back down.

The role and importance of organizational distance was revealed in circumstances when it was found to conflict with subjective distance. The latter reflecting the subject's "mental model" of item similarity. When there was conflict or disagreement between these two, performance suffered, thereby highlighting the importance of a correspondence between the two as pointed out by Roske-Hofstrand and Paap (1986).

Still, maintaining congruence is not a trivial requirement for a number of reasons: different pilots will clearly have different mental models, and a single pilot may have different needs for different organizations over time (i.e., during different phases of flight). Whether this flexibility is best met by time-sensitive changes in organizational structure, by increased navigational tools (i.e., direct access), or by other means, remains the topic of future research.

ACKNOWLEDGMENTS

The authors would like to acknowledge the role of Tom Davis, in contributing the expertise to develop the MFD data base used in these studies. This research was supported by Grant NASA NCC 2-632 from NASA Ames Research Center. Kevin Corker was the scientific monitor.

REFERENCES

- Curran, M. (1991). Electronic library system. SAE Technical Paper 10407.
- Norman, K. (1991). The psychology of menu selection. Hillsdale, NJ: Erlbaum.

- Snowberry, K., Parkinson, S.R., & Sisson, N. (1983). Computer display menus. Ergonomics, 26(7), 699-712.
- Reising, J.M., & Curry, D.G. (1987). A comparison of voice and multifunction controls: Logic design is the key. Ergonomics, 17.
- Roske-Hofstrand, R.J., & Paap, K.R. (1986). Cognitive networks as a guide to menu organization: An application in the automated cockpit. Ergonomics, 29(11), 1301-1311.
- Seidler, K., Wickens, C.D., & Davis, T., Jr. (1991). Distance and organization in multifunction displays (ARL-91-9/NASA A³I-91-3). Savoy, IL: University of Illinois, Aviation Research Laboratory.

Visual Field Versus Muscle Synergy Effects on Control/Display Compatibility: Seeing Is Believing

Dennis B. Beringer
Human Factors Research Laboratory
New Mexico State University

Charles J. Worringham
Motor Behavior Laboratory
The University of Michigan

Abstract

Three components were hypothesized which could affect operator response in the manual control of a system. These included muscle synergy compatibility (MS), geographic control/display compatibility (CD), and visual-field compatibility (VF). A 2x2x2 between-subjects factorial design was used to evaluate all combinations of compatible and incompatible arrangements based upon these variables, using 64 undergraduate students as participants. A static-target acquisition task was used in which subjects manipulated a specially designed joystick to move a cursor to defined target locations. Performance measures included reaction time, movement time, homing time, and frequency and magnitude of directional reversals. Results indicated that visual field (VF) compatibility/incompatibility significantly influenced reaction time, homing time, and reversal frequency and magnitude, while CD and MS manipulations had no significant main effects on performance. Significant gender effects were also found. The results of this study suggest that compatibility of control input and system response is judged primarily by direction of movement in the virtual visual field (self reference). This has implications for the design of systems such as mobile cranes where the operator may be repositioned relative to a fixed directional control.

There are a number of systems employing manual controls in which the operator may assume a variety of orientations relative to the control input device. In some cases the control device is fixed relative to the system it operates while in others the control hardware is more loosely connected, often by flexible cable or radio link. Each of these configurations represents a case where some form of incompatibility between control input and system response may be introduced, either by relative relocation of the operator relative to the control in the first case or reorientation of the control relative to the system in the second case. The latter case is certainly the most recognized as producing input errors (Loveless, 1962) and has, as such, been the source of serious accidents involving small truck-bed-mounted mobile cranes. These accidents have ranged from swinging loads of wallboard through already constructed walls to contacting power lines with subsequent electrocution of the operator.

The case of reorientation of the operator, though less visible, also produces some potential disruption of compatibility and was first recognized by Humphries and Shepard (Humphries, 1958). Their work indicated that simple direction-of-motion compatibility did not always produce the best performance. Although mention was subsequently made by others that location and orientation of the operator should be taken into consideration, little attention was given to these factors. More attention, in fact, has been given to orientation of the control relative to the "display" or effector (Ellis, Tyler, Kim, and Stark, 1991).

A recent attempt was made by Worringham and Beringer (1980) to explore the operator orientation question more thoroughly. This study examined the influences of three forms of directional compatibility: Visual-motor (VM) compatibility in which movement on the display corresponds with movement of the control in the operator's virtual visual field when the operator looks at the control rather than the display; control-display (CD) compatibility which was the conventional "geographic" direction-of-motion compatibility; and visual-trunk (VT) compatibility in which the control movement was in the same direction relative to the operator's trunk as the movement of the display in the visual field. Although VM compatibility ap-

peared to produce performance superior to that obtained with either CD or VT compatibility, the obtained data did not allow one to distinguish between two alternative explanations for its effectiveness. Was this a case of congruence between display motion in the actual visual field and control motion in "virtual" visual field (what one sees when looking at the control when it is not in the same direction from the operator as the display), or was this a correspondence between the display motion and the direction of limb motion relative to the most distal stationary limb segment? In other words, did the motor programming associated with the most usual muscle responses required to operate the control in the expected fashion affect the timeliness and success of the response? This factor, termed muscle synergy (MS), was a potential contributor which could not be ruled out as a factor. Thus, a subsequent study was designed to examine the role played by muscle synergy.

Method

Experimental Design

A 2x2x2 between-subjects factorial design was used to manipulate the independent variables which included geographic control-display (CD) compatibility, virtual visual field (VF) compatibility, and muscle synergy (MS) compatibility. Each variable could take one of two values; compatible or incompatible configuration. Compatibility/incompatibility was defined as follows: CD was defined as compatible when the geographic direction of motion of the control matched that of the displayed element. MS was defined as compatible when flexion of the wrist (right hand used to manipulate the control) produced a leftward movement of the displayed cursor. VF was defined as compatible when the movement of the control in the momentary (virtual) visual field, regardless of control geographic location, matched movement of the displayed cursor in the momentary visual field when the operator's gaze was redirected to the display. Figure 1 details the experimental conditions and the corresponding participant trunk/arm positions, hand positions, joystick positions/movements and cursor movements. It can be seen in this figure that muscle synergy was reversed by inverting the hand, thus causing activations normally requiring flexion to require extension. The design was originally divided into two half replicates, one to be collected at each laboratory site. Each laboratory subsequently collected the full factorial design. The data reported herein is that collected at NMSU.

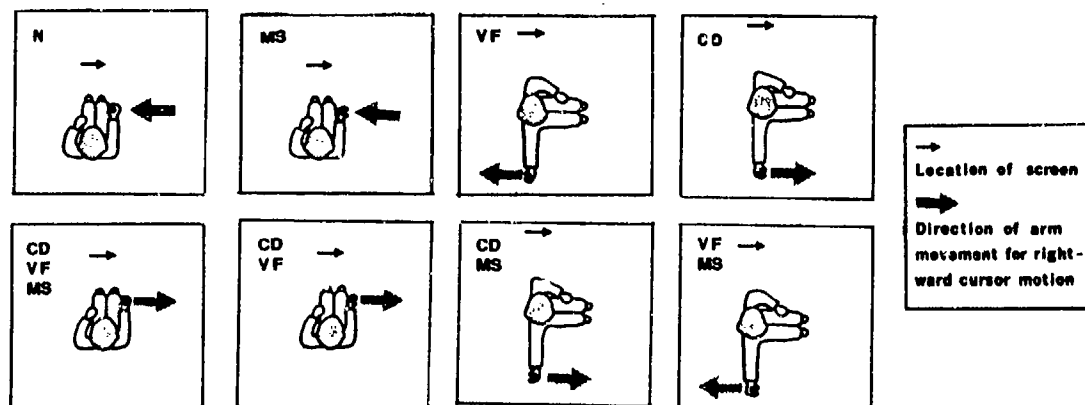


Figure 1. The eight experimental conditions showing hand/arm position, orientation of the control device, and types of compatibility achieved.

Subjects

Thirty-two male and thirty-two female undergraduates ranging in age from 18 to 46 years participated in the experiment. These individuals were students in introductory psychology and were given one-half hour experimental credit for participation. Only three of the 64 individuals stated a left-hand preference.

Apparatus

The control used was a specially designed joystick mounted on a chair having a fully adjustable armrest (fore/aft and height adjustable). Joystick movement was limited to the horizontal plane in one axis, that movement being radial about a vertical axis extending through the operator's wrist. This design was employed to limit and localize the muscle groups used, reducing biomechanical coupling effects. A COMPAQ Deskpro 80286-based microcomputer was used to generate the stimuli on a 14-inch flat-tension-mask color monitor. Reading of joystick input and collecting of data were performed by this system. Viewing distance to the display was approximately 1.5 meters.

Tasks, Stimuli and Performance Measures

The experimental task required the participant to move the displayed system position cursor into a red target ellipse 1 cm wide by 1.2 cm high, as rapidly as possible after both stimuli appeared simultaneously on the screen. Target positions were randomly sampled without replacement for each block of trials from 8 possible locations, four to the left and four to the right of center, at distances of 1.75 cm, 3 cm, 4.5 cm and 6 cm. Successful target acquisition was defined as keeping the cursor within the target ellipse for 300ms and success was signaled to the participant by changing ellipse color to blue. Performance measures included reaction time (RT), movement time (MT; time to first contact with the target), homing time (HT; time to final entry into the target field), and number and magnitude of initial directional reversals.

A psychomotor pretest was administered to each participant to screen for factors that might adversely affect task performance. The pretest consisted of a one-minute tracking task requiring the participant to use a rotary control knob to track a .6 cm circular target moving vertically on the display. Target motion was produced by a summed sinusoid function. This 1-D tracking task was considered adequate for assessing fundamental psychomotor skill while presenting little potential for substantial differential transfer to any of the experimental conditions. Specifically, method of control input was rotary, not linear, and display element motion in pretest was orthogonal to that in the factorial conditions.

Procedure

Participants were assigned to conditions, upon arrival, so that each of the eight conditions contained an equal number of samples after each additional 8 individuals had participated. Each participant was placed in the joystick chair and shown the arm/thumb position for the condition, and then assumed the appropriate trunk/arm/hand position, the chair being adjusted for height and arm length. This adjustment was used to provide comfort for the participant and to control body position variables. The horizontal movement of the joystick was then demonstrated and the participant was instructed to use only wrist motion to achieve the necessary movement. When this preparation was complete, the participant was placed in another chair equally distant from the display and given written instructions for both the pretest and the experimental task. Upon completing the reading of the instructions the participant was given the rotary control for the pretest and instructed in the use of the control. Following verbal review of the written instructions, the pretest was conducted.

Upon completion of the pretest the participant was moved to the joystick chair and given a verbal review of the instructions for the joystick task, after which the joystick task was initiated. The participant then completed four blocks of eight trials each. At the end of the first 4 blocks, a one-minute rest period was allowed, during which the participant was instructed that the conditions of the second part of the task were the same as the first half. When the rest period was completed the participant performed three more blocks of trials. At the completion of this performance assessment the participant was then asked to fill out a questionnaire which

assessed age, handed preference, gender, previous experience or training in selected manual tasks and the amount and location of any fatigue or discomfort.

Results

Data were reduced by averaging across the 8 trials per block, producing mean scores for reaction time (RT), movement time (MT), homing time (HT), and magnitude and frequency of initial directional reversals. These data (logs of RT, MT, HT) were subsequently submitted to analysis of variance (ANOVA) using VF, CD, MS, Blocks and Gender as factors. VF was found to have a significant effect on log RT [$F(1,48) = 5.35$; $p = .0251$] and on log HT [$F(1,48) = 12.53$; $p = .0009$]. The only significant effect on log MT was Blocks [$F(6,288) = 21.561$; $p = .0001$] and this was a clear learning effect across blocks that reached asymptote in block 3 of 7. The main VF effect on MT was in the expected direction but did not attain significance in this sample. Analysis of the Michigan data did find a significant VF effect on MT and this was also true of the pooled analysis. It should also be noted that block effects similar to those found for MT were also found for log RT and log HT. VF produced the only main effect for both magnitude and frequency of reversals ($p = .0001$ in both cases). In the VF-compatible conditions participants were making one reversal per 78 trials (.103 per 8 trials) whereas individuals in the incompatible conditions averaged one reversal for every 5 trials (1.536 per 8 trials).

Gender differences were found for both RT and HT, both favoring the male participants by 100 and 75msec respectively. The significant interaction of VF with Blocks for HT [$F(6,288) = 3.725$; $p = .0014$] indicated a differential rate of performance improvement for the compatible and incompatible conditions. The incompatible groups improved much more from block 1 to block 2 than the compatible groups, but never obtained the level of performance found in the compatible groups. Main effects are summarized in Table 1 and block effects are summarized in Table 2.

Table 1. Summary of means for significant main effects for VF and Gender.

		<u>RT(msec)</u>	<u>HT(msec)</u>	<u>Reversal Frequency</u>	<u>Reversal Mag.</u>
VF	Compatible	556	356	1/78	152
	Incompatible	634	446	1/5	1418
GENDER	Male	547	360		
	Female	644	435		

Table 2. Summary of means for significant Block effects for RT, MT, and HT.

	<u>BLOCK</u>						
<u>Variable</u>	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5</u>	<u>6</u>	<u>7</u>
RT(ms)	685	604	572	577	592	571	562
MT(ms)	811	709	637	640	627	608	624
HT(ms)	535	377	372	395	392	375	360

Discussion

The consistent finding of a VF main effect and the absence of any other effects for CD and MS, be they main or interaction, suggests that the explanation of the effect in present context is solely one related to the virtual visual field. This is, essentially, what one can call a *self-reference* effect. Thus, decisions about direction of movement appear to be based solely upon whether the resulting control movement will be consistent, in the virtual or "redirected" visual

field, with the desired display element movement in the "actual" visual field (view of the display). Variations in muscle synergy appear to have virtually no effect on the time required to perform the acquisition task. These findings are consistent with those of Ladavas and Moscovitch (1984).

It was expected that learning would occur but that performance would stabilize quickly. This was supported by the performance curves obtained across blocks. It is worth repeating that although the VF-incompatible groups improved much more from Block 1 to Block 2 than the VF-compatible groups, the former never quite caught up and showed a relatively stable difference of about 75msec across subsequent blocks. Thus, the effect of disturbed VF compatibility appears to persist and is not easily ameliorated by practice. The same persistence (and significant interaction) appeared for reversals, with a notable reappearance in the first block of trials after the rest period. This suggests a stable difference in performance, both for task execution time and for frequency of errors, despite practice with the system.

Implications for the design and use of systems where the operator is allowed to relocate relative to the manual control element are clear. Relocation to a position between a fixed controller and the display or effector will produce worse performance than that obtained when viewing the control and display/effector simultaneously in the same visual field, given direction-of-motion compatibility, because the control-movement image in the "virtual" visual field is now reversed and no longer congruent with the display/effector movement in the "actual" visual field. In some systems one may wish to make access from this "opposite" side of the control station impossible, forcing a consistent relationship to be observed. There are cases where the need to manage suspended loads manually (usually involving a single operator performing the complete task) requires the operator to change location/orientation and have access from other positions relative to the controls. In these cases one might want to use a device that maintains its orientation relative to the operator so that self reference is consistent relative to both the controller and the effector.

Acknowledgments

The authors would like to thank Jim Kenyon at the University of Michigan for his tireless development of the software required to conduct this study and George L. Short at New Mexico State University for his contributions to hardware development, persistent data collection and reduction activities and contributions to the first draft of the manuscript.

References

- Ellis, S. R., Tyler, M., Kim, W. S. and Stark, L. (1991). Three dimensional tracking with misalignment between display and control axes. Proceedings of the SAE International conference on Environmental Systems, July 15-18, San Francisco.
- Humphries, M. (1958). Performance as a function of control-display relations, positions of the operator and locations of the control. Journal of Applied Psychology, 42, 311-316.
- Ladavas, E. and Moscovitch, M. (1984). Must egocentric and environmental frames of reference be aligned to produce spatial S-R compatibility effects? Journal of Experimental Psychology: Human Perception and Performance, 10, 205-215.
- Loveless, M. E. (1962). Direction-of-motion stereotypes: A review. Ergonomics, 5, 357-384.
- Worringham, C. J. and Beringer, D. B. (1989). Operator orientation and compatibility in visual-motor task performance. Ergonomics, 32(4), 387-399.



Aircraft Experiences During the Vietnam Conflict:

Combat, Cohesion, and Leadership

John A. Ballard, Ph.D.
Wittenberg University
Major Antone E. Gajeski
99th Airlift Squadron

Abstract

During the past 50 years behavioral scientists have published little concerning the combat experience of aircrews. To investigate social-psychological factors in the combat experience of aircrews, a structured interview was developed and used to interview twenty combat aircrew veterans of the Vietnam Conflict. Findings concerning combat, cohesion, and leadership are reported.

The empirical literature on the combat behavior of aircrews is surprisingly small. During World War I flying was new and research focused on physical effects, e.g., lack of oxygen, vibration, and prolonged flight (Dearnaley & Warr, 1948). During World War II Grinkler and Spiegel (1945/1979) studied hundreds of airmen accomplishing a variety of flying duties in the European theater. Their study, Men Under Stress, remains the largest, most indepth psychological study of combat aircrew behavior.

Grinkler and Spiegel were the first to report that fliers exhibited a higher degree of group cohesion than ground forces. Likewise they documented the profound effect of the combat leader on morale. Lentz (1986) also emphasized the role of leadership and group cohesion in the behavior of WWII aircrews. Forty years after the war, Lentz interviewed WWII commanders of the 56th Fighter Group. He found vivid recollections of the combat experience. For example, he shared Colonel Mahurin's comments on group cohesion:

This really came into play the first time I saw enemy aircraft attacking our bomber formations. All I could think was that, "they are hurting our boys" and I instantly went to attack without thinking of anything else. Up to then I had really been scared silly to go into combat, but the peer pressure kept me from quitting ... When I saw the Germans for the first time, it never occurred to me to be scared. I just wanted to get even with them. (Lentz, 1986, p. 37)

Janus (1949) also reported studies of WWII combat aircrews. Part of The American Soldier, the classic work by Samuel Stouffer and his associates on the combat behavior of WWII ground forces, Janus described the differences between the combat environment of the infantryman--sustained danger in a deprived environment--and that of the airman--short duration encounters with ground time in a relatively safe environment. Janus noted that the fixed tour length of the flier offered a goal which benefitted the psychological health of combat crews.

The views expressed herein are those of the authors and do not necessarily reflect the views of the United States Air Force or the Department of Defense.

It has been nearly 50 years since the last major studies of the behavior of aircrews in combat. Our review found only one study from the Korean Conflict. Ruch (1953) interviewed 75 fliers in the Far East Air Force. The results however were reported only as behavioral anecdotes illustrating effective and ineffective behaviors. Likewise we found only one study from the Vietnam Conflict. Kantor, Klinefelter, and McFarlane (1978) interviewed 16 members of the Red River Valley Fighter Pilots Association, pilots who had flown into the most heavily defended areas of North Vietnam. These fliers reported they experienced more stress enroute to targets than over the targets. More recently Jones (1985) discussed some of the effects of combat on fliers, including the problem of fatigue, the importance of leadership, and possible mechanisms by which airmen cope with the stresses of combat.

The findings reported here were part of an exploratory descriptive study of combat behavior of aircrews who flew and fought in the Vietnam Conflict. We defined combat behavior as "individual and group behavior occurring in a combat environment--an environment in which hostilities have occurred, are occurring, or in which the individual perceives that hostilities are probable" (Ballard, 1988, p. 200). The overall study investigated personal, group, and situational factors affecting aircrew combat behavior. Reported here are some of the findings concerning the experience of combat, group cohesion, and leadership.

Method

Research Instrument

A structured interview containing 47 questions was developed for this study. The senior author's course on combat behavior at the Air Force Institute of Technology provided an overview of the social-psychological literature and identification of factors to be explored in this pilot study. Questions were grouped as follows: flying experience prior to arriving in SEA, introduction to combat, day-to-day life in theater, flying in combat, and the physical effects of flying. Findings reported here were developed from responses to 10 of the open-ended questions. [Copies of the interview questionnaire may be obtained from the senior author.]

Participants

Twenty combat veterans volunteered to participate anonymously in this study. Participants were recruited by word of mouth, from USAF Academy graduates, and from Daedalion members. Fifteen were on active duty; five were retired. The study included 17 pilots (4, F-100; 2, F-105; 3, C-130; 2, F-4; and 1 each, B-57, B-52, EC-47, UH-1, A1-E, O1-E), an EWO (F-4), a WSO (RF-4), and a tail gunner (B-52). Seventeen of the interviews were conducted by the second author, a USAF pilot and USAF Academy graduate. The other three were conducted by assistants of the senior author.

Procedure

Most of the interviews were accomplished in the interviewee's work setting. Permission was obtained to audio tape the interviews. To facilitate rapport, the interviewer was in Air Force uniform. Transcriptions of the interviews were then prepared for analysis.

Results

Prior to Combat

The participants covered most of the time span of the American experience in SEA, from 1965 to 1973. Most were stationed in South Vietnam and Thailand. The median age was 24 and half were captains. The missions flown were as diverse as their aircraft. Thirty percent of the missions flown were at night with the concurrent problems of flashblindness, vertigo, or disorientation. Sixteen of the 20 airmen reported that they thought about combat while still in flying training, one noting he thought about it "constantly". Sixteen were also volunteers for SEA. Six volunteered to prove their skills and apply their training. Five saw it as a logical career development. The majority learned something about combat prior to their departure for SEA. Primary sources of information included instructor pilots and others who had returned from SEA.

Exposure to Combat

All of the airmen reported exposure to combat. Eleven said they were exposed to enemy fire very often and that their missions were highly dangerous. Many fliers mentioned that they were fired on much more than they realized, but they had been too busy to notice. Others were not aware of being fired upon until they were hit. Engine failure seemed to be a bigger concern than combat to some crewmen of single-engine aircraft.

Prior to their first mission, most reported that they had been nervous, anxious or excited. Most received orientation missions to ease them into the theater of operations.

How did the aircrews react to their first experience of combat? Most felt surprised. There was someone actually firing at them. For some it was exhilarating: "it was an intense high", "awesome". For others there was fear. One pilot reported flinching and ducking behind the glareshield of the aircraft as if to escape the flack around him. Three fighter pilots reported an attitude of "if you're gonna shoot at me, I'm gonna get you". For aircrew of the unarmed aircraft there was more a sense of shock that there was an enemy out there. Six airmen commented on the awesome spectacle of gunfire: "waterfalls of tracers, firehoses of bullets, and pretty Fourth of July fireworks". Although many of the fliers were scared when they first encountered combat, all stayed too busy to think about what was happening. Usually the full impact of the encounter was not felt until they were on their way home.

It is interesting to note that two airmen compared their airborne combat experiences to combat experiences they experienced on the ground. Mortar attacks sometimes forced them to take shelter. Unable to shoot back or flee as in the air, both felt helpless.

Cohesion and Leadership

As the literature review noted, group cohesion is a vital ingredient in flying units. This study found perceived cohesion to vary among units. Eleven airmen rated their unit's cohesiveness as high; five, moderate; and three, low. One worked alone. The organizations with the highest cohesion were the

units where aircraft were deployed in formation in a team toward achieving some objective. These squadrons seemed to experience the thrills and disasters of combat as units instead of just as individuals. Flights within fighter squadrons were the most cohesive. They fought together and "watched out for each other".

Bad leadership was seen as affecting morale. Several airmen described how morale and cohesion changed as a function of the leader. A poor commander's decisions and example could destroy the squadron. When asked what qualities and behavior make a good combat leader, sixteen replied that the good leader sets the example, knows his job, and has confidence in himself. He is a competent flier and frequently takes part in the missions. The most frequently cited example was General Robin Olds.

Discussion

For most individuals who have known combat, the experiences of combat remain vivid in memory. The literature on combat behavior is replete with testimonies about how "I never felt more alive" and comments on the spectacle of war. We suggest that in spite of the limitations of retrospective data reported 15 to 20 years after the event, there is much to be learned from such interviews. The findings reported here in a small convenience sample of limited generalizability suggest that the social-psychological experience of the combat airmen and the combat groundman have commonalities that are simply part of the human experience of combat. Both experience fear primarily before and after combat but usually not during combat. The initial response of both airmen and the infantrymen to that first experience of combat is that someone is shooting at them, a disbelief, an awakening into the reality of war. Likewise, both need a sense of empowerment. Regardless of the effectiveness of the weapon at hand, a weapon facilitates the coping and allows for a feeling of fighting back. Both speak of the phenomenological aspects of combat, the spectacle, even the beauty. For both, taking care of one's buddies is a foremost motivator. For both, cohesion must be coupled with good leadership. For 50 years there had been little research about the combat behavior of aircrews. There is much meaningful research that can be done.

References

- Ballard, J.A. (1988). The study of combat behavior: Who will pursue the beast in the jungle. Proceedings of the Eleventh Symposium of Psychology in the DOD, 11, 199-203.
- Dearnaley, E.J. & Warr, P.B. (1948). Aircrew stress in wartime operations. New York: Academic Press.
- Grinkler R.R. & Spiegel, J.P. (1979). Men under stress (3rd ed.). New York: Irving. (Original work published 1945)
- Janis, I.L. (1949). Morale attitudes of combat flying personnel in the air corps. In S.A. Stouffer, A.A. Lumsdaine, M.H. Lumsdaine, R.M. Williams Jr., M.B. Smith, I.L. Janis, S.A. Star, & L.S. Cotrell Jr. (Eds.), Perspectives in social inquiry: Vol. 2. Studies in social psychology in World War II.

Jones, D.R. (1985). U.S. Air Force Combat Psychiatry (Report No. USAFSAM-TR-85-83). Brooks AFB, TX: USAF School of Aerospace Medicine.

Kantor, J.E., Klinefelter, L., & McFarlane, T.A. (1978). Methodology to assess psychological stress and its impact in the air combat environment (Report No. AD-A0053474). Brooks AFB, TX: Air Force Human Resources Laboratory.

Lentz, T.L. (1986). Combat leadership: 56th Fighter Group 1943-1944 (Report No. AD-A168864). Maxwell AFB, AL: Air Command and Staff College.

Ruch, F.L. (1953). Incidents of leadership in combat (Report No. AF 18(600)-468). Maxwell AFB, AL: Human Resources Research Institute.

AD-P006 940


Aircrew Coordination: What Does it Take?¹

Rene'e Stout²
Carolyn Prince
David P. Baker
Maureen L. Bergondy
Eduardo Salas
Naval Training Systems Center

92-17673



Abstract

The current investigation compared the perceptions of the importance of aircrew coordination skills for four military aviation communities. Results revealed that all of the communities investigated consider aircrew coordination skills to be important. However, the relative importance of aircrew coordination skills was found to differ by community.

Over the past twelve years, aviation has seen a surge in the use of Aircrew Coordination Training (ACT) programs. These programs seek to train aircrews to effectively interact and coordinate their activities, with the expectation that this will help eliminate some of the human error that occurs in the cockpit. Interest in the development and improvement of ACT continues based on the important relationship between coordination among the crew and mission effectiveness and safety.

Despite widespread implementation of ACT, a review of a number of these programs (Jensen, 1987) showed differences in the training content which highlights an existing lack of consensus about what constitutes ACT, and what skill areas need to be trained. As noted by several researchers (Franz, Prince, Cannon-Bowers, & Salas, 1990; Prince & Salas, in press), diversity in these programs persists partially due to a lack of research on the skills and behaviors that should be included in such training.

Prince and Salas (1989) identified skill behaviors taught in a number of ACT programs and classified them into seven skill areas (i.e., Decision Making, Assertiveness, Mission Analysis, Communication, Leadership, Adaptability/Flexibility, Situational Awareness). Building upon this research, Franz et al. (1990) then described a methodology they used to identify the behaviors that underlie effective coordination for a single aviation community. Initially, interviews were conducted with pilots to

¹The views herein are those of the authors and do not reflect the official position of the organizations with which they are affiliated.

²This research was supported in part by an appointment to the Research Participation Program at the Naval Training Systems Center administered through Oak Ridge Associated Universities through an Interagency agreement between the U.S. Department of Energy and NTSC.

identify critical incidents and behaviors that related to aircrew coordination. Subsequently, a short questionnaire was developed that included critical behaviors derived from the interviews. This was given to aircrew members who rated each behavior on criticality, frequency, and importance of training. Last, the behaviors were clustered into the seven skill dimensions identified by Prince and Salas (1989).

While Franz et al. (1990) provided some preliminary insight regarding the behavioral skill dimensions associated with ACT in military helicopter crews, questions remain regarding the generalizability of the skills to aircrews in other communities. Such information would be valuable with respect to understanding crew coordination demands, and help in the design and development of ACT programs. The present study attempted to investigate this issue by replicating the methodology described by Franz et al. (1990), across several different military communities.

Method

Participants

Naval aviators from the MH-53, F-14, A-6, and T-44 communities participated in the study. Overall, the sample consisted of 35 MH-53 pilots, 51 F-14 crew members, 37 A-6 crew members, and 122 T-44 pilots. Of the F-14 and A-6 respondents, 49 pilots and 39 Naval Flight Officers (NFOs) were included. Respondents for all other platforms were pilots. Respondents ranged in rank from Ensign to Commander. The average age was 28, average years in service was six, and the operational flight hour average was 1090 (The undergraduate aviator population was excluded from the last calculation). In the four communities, there were 107 aviators who were either instructors or Fleet aviators. The remaining participants were students or were aviators in a training program to become qualified for that particular aircraft.

Procedure

This investigation employed a revised version of the questionnaire described by Franz et al. (1990). The questionnaire was composed of a series of behaviors related to teamwork in the cockpit and included such behaviors as, "Cross check information for agreement/confirmation" and "Provide assistance to crew members as needed". For the present study, items were rewritten to eliminate redundancy, and the form was tailored to each specific platform by Subject Matter Experts (SMEs) from each community. Although the wording on the questionnaire varied from one community to the next, the items were considered examples of the skill dimensions under investigation; therefore, variation in questionnaire length and wording was not considered a problem. Items on each of the questionnaires were categorized into the seven behavioral dimensions identified by Prince and Salas (1989) by six research psychologists who were experts in the area of ACT.

Participants rated each item on a 7-point scale with respect

to its criticality, difficulty, relative time spent, difficulty of learning, importance to training, and overall importance. These scales were employed based on their common use in the job analysis literature (Sanchez & Levine, 1989). Dimension scores were computed for each scale by averaging the ratings for all of the items categorized under that dimension for a given scale. This procedure resulted in an average rating on each scale (e.g., criticality) for each of the seven aircrew coordination skill dimensions (e.g., communication).

Results

Results are only reported for the overall importance ratings in order to provide global information on the relative importance of the seven skill dimensions across the four aviation communities.

Data were first analyzed to determine whether differences in ratings varied as a result of whether respondents were pilots or NFOs; Fleet aviators, instructor aviators, or aviators in training (including student T-44 aviators and aviators in a training program to become qualified for each of the other aircraft types); or as a function of rank. Thus, a series of Analyses of Variance (ANOVAs) were performed to determine if these subgroups varied with respect to their overall importance ratings on each of the seven skill dimensions. These analyses produced no significant effects.

A Multivariate Analysis of Variance (MANOVA) was conducted to examine potential differences in the MH-53, F-14, A-6, and T-44 communities on the overall importance of the seven aircrew coordination skills. This analysis produced a significant result using Wilk's criterion, $F(3,199)=3.47$, $p \leq .05$. Furthermore, subsequent univariate ANOVAs for each dimension were also found to be significant: Decision Making, $F(3,200)=4.94$, $p \leq .003$; Assertiveness, $F(3,215)=14.55$, $p \leq .000$; Mission Analysis, $F(3,214)=7.89$, $p \leq .000$; Communication, $F(3,207)=10.42$, $p \leq .000$; Leadership, $F(3,199)=4.97$, $p \leq .003$; Adaptability/Flexibility, $F(3,215)=6.01$, $p \leq .001$; and Situational Awareness, $F(3,215)=4.70$, $p \leq .004$.

Table 1 reports the mean overall rating for each of the seven skill dimensions by aviation community. Tukey tests were performed to determine which community rated each dimension significantly different for the overall importance scale. These comparisons indicated that the MH-53 community rated each dimension as significantly more important than the other communities, with the exception of situational awareness in the A-6. In addition, the means listed in Table 1 indicate that all dimensions were rated as important by each of the communities, with the lowest rating obtained still being on the high end of the rating scale.

Table 1

Mean overall ratings by community for each skill dimension

Dimension	Aviation Community			
	MH-53	F-14	A-6	T-44
Decision Making	5.62	4.81	4.77	5.06
Assertiveness	6.05	5.06	4.94	4.83
Mission Analysis	5.87	5.06	5.06	4.88
Communication	5.94	5.09	5.10	5.00
Leadership	5.49	4.65	4.68	4.89
Adaptability/Flexibility	5.93	5.10	5.08	5.32
Situational Awareness	5.71	5.10	5.12	5.02

Discussion

The most striking result of this research is the finding that aviators from three diverse fixed wing communities did not significantly differ in their ratings of any skill dimensions for overall importance to aircrew coordination. The differences in the equipment flown, the missions, and the experience of the aviators did not affect their perceptions of the importance of aircrew coordination skill behaviors.

Aviators from the only rotary wing community included in this analysis, the MH-53, rated most of the dimensions significantly higher for overall importance to aircrew coordination than did fixed wing aviators. Only the dimension of situational awareness showed no significant difference in the ratings by MH-53 and A-6 aviators. Since the MH-53 community was the only rotary wing community included in this analysis, it cannot be determined if the differences in ratings noted were due to the type of aircraft, the missions flown, the size of the crew (theirs was the only crew larger than two people), or training within their community. Further analyses will include additional rotary wing communities to see if the MH-53 responses can be replicated.

In conclusion, it is interesting to note both similarities and differences in ratings of the importance of aircrew coordination skills for aviators from various operational communities. The similarities suggest that the skills identified by Franz et al. (1990) are generic to a certain extent rather than being specific to one type of aviation community (i.e. the community for which it was developed). Moreover, since the dimensions were found to be important by all of the communities, their relevance to each community is supported. The results suggest that ACT programs should include content relating to the seven skill dimensions and can be general in that respect. The

differences in ratings suggest that the development of an ACT program for a given community must contain a degree of specificity as well. This information should be used to guide curriculum development at each of the four communities surveyed. The current investigation was a first step in determining whether overall differences occurred. Further research must be done to determine the effectiveness of training the aircrew coordination skills.

References

- Franz, T. M., Prince, C., Cannon-Bowers, J. A., & Salas, E. (1990). The identification of aircrew coordination skills. Proceedings of the 12th Annual Department of Defense Symposium, 97-101.
- Jensen, R. S. (1987). Cockpit resource management training: Final Report (Contract No. DTFA01-80-C-10080). Arlington, VA: Systems Control Technology, Inc.
- Prince, C. W., & Salas, E. (1989). Aircrew performance coordination and skill development. In D. E. Daniels, E. Salas, & D. M. Kotick (Eds.), Independent research and independent exploratory development (IR/IED) programs: Annual report FY88 (Technical Report No. NTSC 89-009). Orlando, FL: Naval Training Systems Center.
- Prince, C. W., & Salas, E. (in press). Training for teamwork in the military aircrew: Military CRM programs and research. In E. L. Weiner, B. G. Kanki, and R. L. Helmreich (Eds.), Cockpit resource management. Orlando, FL: Academic Press.
- Sanchez, J. I., Levine, E. L. (1989). Determining important tasks within jobs: A policy-capturing approach. Journal of Applied Psychology, 74(2), 336-342.



Attention and Skill

92-17674



Anthony J. Aretz, Ph.D.
US Air Force Academy

Abstract

In this paper I propose a theoretical framework for the integration of attention and skill into a single unified theory. The main tenant of this framework is that the primary purpose of human cognition is to select constraints to reduce task specific behavioral alternatives. Attention is required for constraint selection when there is no skill for a task. When a task specific skill develops, attention is no longer needed. Hence, attention and skill are at the opposite ends of a human performance continuum.

What is it that allows effortless skilled performance on some tasks, but requires effort on others? This is a difficult question and psychology has yet to arrive at a widely accepted explanation. In this paper, I strive to provide a comprehensive framework that can be used to explore some of the fundamental psychological issues of human performance. The framework centers on task constraints as the basis of human performance and draws heavily on both the ecological (e.g., Gibson, 1979) and cognitive approaches (e.g., Lindsay and Norman, 1977) to human performance.

Task Constraints

The task constraint concept is derived initially from the study of kinetics. In kinetics, constraints limit a moving body to a particular trajectory. For example, the trajectory of a projectile is constrained by its initial launch conditions and gravity. The same concept can be applied to human behavior. Goal directed behavior can be viewed as emerging from the interaction between task constraints and the behavioral alternatives (i.e., degrees of freedom) present in a task.

Task constraints are found either in the environment or mind and control action through perception and attention. The main thesis of this paper is that perception and attention act together to constrain behavior. A skill provides the perceptual ability to exploit environmental constraints and attention provides any remaining cognitive constraints that may be required. Accordingly, the main difference between skill and attentionally constrained behavior is the source of the constraints. In a skill, perception constrains behavior through the direct perception of environmental information. In attention, cognitive processing is used to constrain behavior. Thus, attention is required when a task cannot be uniquely constrained by skilled perception.¹

Direct Perception

The idea that environmental information can serve as a source of behavioral constraints is not usually explicitly incorporated into a traditional information processing approach to human performance (e.g., Neisser, 1967; Lindsay and Norman, 1977; Wickens, 1984). The information processing approach assumes that information in the environment is impoverished and does not contain sufficient information to provide task constraints. Gibson's theory of direct perception (1979; see also Michaels and Carello, 1981) suggests that the environment contains meaningful information that can uniquely specify behavior without further elaboration through cognitive processing.

For example, skilled drivers can keep a vehicle on a road through the direct perception of the road shape, lane markings, traffic signals, and optical flow. This perceptual information can directly

¹For the purposes of this short paper, the discussion centers mostly on visual perception as a source of behavioral constraints and does not address other possible sources (e.g., motor, kinesthetic, or biomechanical).

constrain driving behavior without the need for attentional processing. How many times have you driven to work without realizing how you got there? Or, have you ever driven to work when you really wanted to drive somewhere else? The fact is, you probably weren't paying attention. Perception was constraining your behavior automatically without attention, i.e., you were on automatic pilot.²

Skilled Behavior

This theoretical interpretation rests on the assumption that many other skilled behaviors are also constrained by environmental information. If environmental constraints are the main source of skilled behavior, then skill acquisition is a process of learning to directly perceive environmental constraints for the control of behavior. What differentiates levels of a skill, or one skill from another, is the perceptual sensitivity to the relevant environmental constraints.

The main assumption of this approach is that for a skill to develop, the environmental information used to constrain behavior must be invariant. An invariant has been defined as a pattern of behavior or stimulation that remains constant over space or time under a given set of physical transformations (Kugler and Turvey, 1987). A more simple definition is that there must be consistent mappings within the perception-action cycle. To illustrate, consider a novice musician learning to play the piano. The novice must first learn the correspondence between notes on a manuscript and the appropriate finger positions on the keyboard, which are invariant. Through practice, the novice eventually acquires a skill that constrains behavior by directly perceiving the invariant information in the score, i.e., the notes. Since direct perception is being used to control their playing, behavior can occur without attention. Similar invariant relationships existing in other contexts provide the informational constraints necessary for other skills. For example, many automaticity researchers have shown that skills for simple laboratory tasks can develop and preclude attentional processing (see Logan, 1985 for a review).

Consistent one-to-one perception-action mappings are not the only invariants that can be used in skill development. Higher order invariant mappings can also exist and lead to skill development (Kramer, Strayer, and Buckley, 1990). For example, the novice pianist's skill continues to increase as they learn to perceive the higher level invariants present in music (e.g., chords, progressions, and key signatures). Similar higher order invariant mappings are necessary for continued improvement in other skilled behaviors. It is important to emphasize that the invariant hierarchical organization required for highly skilled behaviors do not develop in the mind or brain as traditionally assumed in an information processing approach. What develops is a unitized perceptual sensitivity to the hierarchical constraints present in the environment.

Unitization is the cognitive process that allows for the development of perceptual sensitivity to hierarchical invariants (LaBerge, 1981). As used in this paper, unitization refers to the consolidation of many single invariant perceptual mappings into a single perceptual chunk. The previously separate mappings drop out of the control process (and hence out of the mind). Through unitization the many one-to-one invariant perception-action mappings that were present early in skill development are replaced by fewer one-to-many mappings as the skill develops. Eventually, a single higher-order perceptual invariant could be used to constrain a whole series of invariant actions. For example, skilled musicians perform a well known score by perceiving sections of the score, not the separate notes. The separate notes were only perceived when the score was unfamiliar, and eventually they were unitized into sections.

The idea that some form of informational chunking underlies skilled performance has been with us for some time. Bryan and Harter (1897) provide one of the earliest descriptions of behavioral chunking. Bryan and Harter collected data on skilled telegraphers and found that as telegraphy skill develops, there are plateaus in performance that are indicative of unitization. Bryan and Harter proposed

²In this paper, I implicitly make a distinction between attentional control and attentional monitoring. Attention can be used to monitor skilled performance without necessarily controlling it.

that these plateaus were the result of new skills acquired for perceiving informational chunks in the auditory telegraphy signal. Each plateau was an indication that single dots and dashes had become letters, letters had become words, or words had become phrases. The chunking hypothesis was also based on the highly stereotyped and less varied data obtained from highly skilled telegraphers. Within the current context, the chunking hypothesis can be explained in terms of developing a perceptual sensitivity to the higher-order invariant constraints present in the telegraphy signal.

It is important to re-emphasize that the information necessary for skilled telegraphy, and many other skills, is already contained in the task. The information is not created in the mind. What develops is a perceptual skill to respond to information in the environment that constrains behavior.

Attentional Constraints

There are three situations where behavior may not be controlled through skilled perception. Behavior may be underconstrained due to a lack of skill, underconstrained by a lack of information available, or overconstrained due to excess or conflicting constraints. In any of these situations, attention will be required to select or reorganize constraints so that behavior can be controlled (see also Neumann, 1984, 1987 for a related treatment).

Lack of Skill

A lack of skill is the inability to exploit the environmental constraints present in a task to control behavior. In these situations, attention is used to control behavior by selecting from behavioral alternatives. As a skill develops, control of the behavior will gradually shift from attention to the perception of the environmental constraints.

The goal of a novice, through practice, is to acquire the perceptual sensitivity to invariant relations that uniquely constrain behavior. For example, a piece of music may seem like a foreign language to a novice since the novice is insensitive to the potentially constraining information present in the music. Hence, performance is underconstrained and attention must be used to compensate for the lack of perceptual constraints. Finger movement must be controlled by attentionally linking perceptual information to action. Through practice, the skilled pianist becomes perceptually sensitive to the music's constraints that control the performance, without attention.

Thus, the goal of skill development is to acquire the perceptual sensitivity to the appropriate environmental constraints that can be used to control behavior. Attention is only used to compensate for the lack of skill by controlling behavior that otherwise would be perceptually constrained by environmental information.

Lack of Invariant Constraints

In contrast to lack of skill, a lack of constraints implies there is a lack of invariants that can be used to control behavior. The mappings that do exist vary and cannot be used as a basis for skill development. For example, no amount of practice will allow a pilot to land an aircraft without attention. Although some invariants are present, such as the optimal glide path, there are variants, mainly wind, that enter the task and prevent a unitized skill from developing. Because of the variants, landing an aircraft is an attention demanding task. The need for attentional constraints could only be eliminated if aircraft landings were completely invariant.

Another similar situation that might arise, is a lack of constraints due to the degradation or absence of constraints. Attention will be required in these situations, even if a skill exists for the task, e.g., haze or fog on a runway during an approach.

Conflicting Constraints

Another problem results from conflicting constraints. In this situation, the information available contains multiple constraints that specify incompatible behaviors. For example, two different constraints may specify two different behaviors that cannot be performed concurrently. The Stroop task provides a good laboratory example of a task with conflicting constraints. The Stroop task requires a subject to say the printed color of a word while ignoring the text of the same word that spells a color (e.g., if the word red were printed in green, the subject would say green). For skilled readers, it is almost impossible to quickly name the ink colors of several words when they conflict with the text. The ink color and word text provide two constraints that lead to mutually exclusive behaviors, creating a rather difficult task to perform.

Two potential solutions, that may occur in sequence, can be used to reduce the conflict among constraints. The first is constraint selection and the second is constraint integration.

Constraint selection occurs in situations when there is an inability to satisfy all potential constraints simultaneously. In order to perform the task, attention must select one constraint to control behavior and ignore the others. However, as in the Stroop task, this may be difficult for a skilled perceiver to ignore conflicting constraints in the same modality. If the constraints were in different modalities, attentional selection would be easier. For example, it is difficult to read an article while conversing with someone on the phone, but attention could select one of the inputs and ignore the other.

Alternatively, a task with conflicting constraints can be made easier by reducing the incompatibility between the constraints. Pritchatt (1968), for example, increased performance on the Stroop task by requiring subjects to press a key corresponding to the color rather than saying the name. There is less conflict between a motor and verbal response. In addition, performance could also be improved by simply integrating the constraints so the text and color specified the same behavior.

Conclusion

In this paper I have suggested that as long as a person is perceptually sensitive to environmental constraints, action can occur without attention through direct perception. If the situation does not provide for environmentally constrained behavior, then attention must control behavior, i.e., when there is a lack of constraints, lack of skill, or an incompatibility among constraints. By proposing this framework, I have tried to bring together two traditionally disparate fields of inquiry, attention and skill (or ecological and cognitive psychology for that matter). In sum, the difference between attentionally controlled and skilled behavior can be formally defined as a difference in the source of the task constraints. Attention is required when a task cannot be uniquely constrained by a skill. I hope this integration helps to unify the theoretical foundation of cognitive psychology.

References

- Bryan, W. L. & Harter, N. (1897). Studies in the physiology and psychology of the telegraphic language. Psychological Review, 1(4), 27-53.
- Gibson, J. J. (1979). The ecological approach to visual perception. Boston: Houghton Mifflin.
- Kramer, A. F., Strayer, D. L., Buckley J. (1990). Development and transfer of automatic processing. Journal of Experimental Psychology: Human Perception and Performance, 16, 505-522.
- Kugler, P. N. & Turvey, M. T. (1987). Information, Natural Law, and the Self-Assembly of Rhythmic Movement. Hillsdale, NJ: Erlbaum.
- LaBerge, D. (1981). Automatic information processing: A review. In J. Long & A. D. Baddeley (Eds.), Attention and Performance IX (pp. 173-186). Hillsdale, NJ: Erlbaum.

- Lindsay, P. H. & Norman, D. A. (1977). Human Information Processing. New York: Academic Press.
- Logan, G. D. (1985). Skill and automaticity: Relations, implications, and future directions. Canadian Journal of Psychology, 39, 367-386.
- Michaels, C. F. & Carelio, C. (1981). Direct Perception. Englewood Cliffs, NJ: Prentice-Hall
- Neumann, O. (1984). Automatic processing: A review of recent findings and a plea for an old theory. In W. Prinz & A. F. Sanders (Eds.), Cognition and Motor Processes (pp. 255-293). Berlin: Springer-Verlag.
- Neumann, O. (1987). Beyond capacity: A functional view of attention. In A. F. Sanders (Ed.), Perspectives on Perception and Action. Hillsdale, NJ: Erlbaum.
- Pritchatt, D. (1968). An investigation into some of the underlying associative processes of the Stroop colour effect. Quarterly Journal of Experimental Psychology, 20, 351-359.
- Wickens, C. D. (1984). Engineering Psychology. Columbus: Merrill Publishing Company.



Virtual Reality and Education

Christopher D. Wickens

University of Illinois, Institute of Aviation, Aviation Research Laboratory,
Savoy, IL
and

DFBL, U.S. Air Force Academy, Colorado Springs, CO

ABSTRACT

In this paper, five components of the virtual reality concept are analytically defined: 3D perspective, dynamic rendering, closed loop interaction, inside-out perspective and enhanced sensory feedback. It is argued on the basis of empirical data from a variety of sources, that those components that improve performance by reducing effort may actually inhibit learning or long term retention. Closed loop interaction in contrast, while not reducing effort, appears to have a beneficial effect on retention. The importance for learning of directing users' attention to the link between the VR perspective and a more artificial perspective is also highlighted.

The concept of virtual reality is created by an impressive, exciting technology which readily engages the interest of the user. As a consequence, it is reasonable that the concept should be considered by educators as a plausible way of exploiting the technology of advanced computers to improve or expand the learning environment. This paper first presents the arguments that may be made for introduction of virtual reality, but then takes a critical look at the constraints on its applicability to education, while drawing on empirical data from simulation, experimental laboratory studies, and basic human factors principles of design.

Virtual Reality: Its Components and Justification

Virtual reality, in fact, is not a unified thing, but can be broken down into a set of five features, any one of which can be present or absent to create a greater sense of reality. These consist of:

1) Three-dimensional (perspective and/or stereoscopic) viewing. Thus, the geography student who views a 3D representation of the environment has a more realistic view than one who views a 2D contour map. 2) Dynamic vs. static display. A video or movie is more real than a series of static frames of the same material. 3) Closed-loop (interactive or learner-centered) vs. open-loop interaction. A more realistic closed loop mode is one in which the learner has control over what aspect of the learning "world" is viewed or visited. That is, the learner is an active navigator as well as an observer. 4) Inside-out (ego-referenced) vs. outside-in (world-referenced) frame-of-reference. Within the dynamic mode (see 2), the more realistic inside-out frame-of-reference is one in which the image of the world on the display is viewed from the perspective of the learner icon's momentary frame-of-reference. Thus, the explorer of a virtual undersea world will view that world from a perspective akin to that of a camera placed on the explorer's head; rather than, for example, a fixed camera pointing north. For

conventional navigational displays, we contrast here north-up vs. track-up maps (Harwood & Wickens, 1991; Aretz, 1991).

5) Enhanced sensory experience. An integral part of many virtual reality systems is the enhanced feedback which might characterize the tactile feedback from a "data glove" as the hand touches a "virtual object," the proprioceptive feedback from motion through a virtual world, or the auditory feedback emanating from aspects of the environment. We can also include here the use of sound to more realistically characterize the aspects of the environment, although this is not "feedback" in the formal sense.

It is easy to see how each "layer" of these five elements of realism can be peeled away or added as desired. In this regard, it is also important to realize that each layer brings with it added costs, or added sources of unreliability in equipment maintenance. As an instructional medium, these costs must be justified in terms of their educational benefit. In this regard, the following justifications may be cited.

1) Motivational value. It may be argued that the intrinsic interest fostered by dynamically interacting with the pictorial learning environment, like a video game, can provide the necessary (but not sufficient) spark of motivation for the student to be involved with the learning environment.

2) Transfer of learning environment. The value of field trips in learning is that they allow the student to directly experience that material they are to be learning. That is, learning takes place in a real-world context, and its relevance to actual world issues can often be more directly seen. Virtual reality may attempt to capture this "field trip-like" experience.

3) Different (novel) perspective. This is an oft-claimed advantage for scientific visualization (a close cousin to virtual reality). The scientist (or other learner) can gain insight from a set of data by exploring it (the interactive component) and viewing it from novel perspectives (the frame-of-reference component; Zorpette, 1989; Cohen, 1989).

4) A "natural" interface. Much like direct manipulation (Hutchins, Hollan, & Norman; 1985), also a close cousin of virtual reality, the argument is made that the direct and natural navigation through a virtual reality environment, because it captures the essence of walking through a real environment, can be accomplished without the cognitive effort required of many less natural computer interfaces (e.g., those based upon symbolic command strings, or verbal/world-referenced status information rather than spatial and ego-referenced information).

Returning to our presentation of the features of VR, we see that they can be roughly placed into two categories--those emphasizing the naturalness of the interface and in so doing reducing effort, and those emphasizing dynamic interaction and novel perspective.

In the remainder of this paper, we argue why it is important to keep this distinction clear as we consider the conditions in which VR can and cannot assist learning, as we argue that those features of an interface that may reduce effort and increase performance may actually reduce retention.

Effort Reduction, Performance, and Learning

Within the human factors community, volumes have been written regarding the concept of mental workload and the desirable design goal of creating systems that do not place unnecessary load on the human operator's limited cognitive resources (e.g., Wickens, 1991). Indeed it is also true that many options in flexible advanced computer systems go unused because of the high cognitive effort required to realize their more powerful benefits. As noted above, effort reduction is also a justification for the choice of natural visual-spatial-manual interaction tools characteristic of the direct manipulation interface and virtual reality.

It is easy to support the claim that the tools for direct manipulation interface are useful if they eliminate the cognitive workload involved in making transformations from intentions to arbitrary key presses, or from a displayed digital information code to an understanding of its meaning (Sweller et al., 1990). However, one must be more cautious in assuming that all features of virtual reality that reduce the cognitive effort of interaction and may therefore improve navigational performance, will be of benefit for long term retention of information. This is particularly true to the extent that such features may eliminate effort-demanding cognitive transformations that are necessary to form mental links between different representations of the material to be learned. Several examples may be cited of data suggesting that the five different features of virtual reality which may support better performance of the operator while interacting with the learning environment, do not necessarily support better long term retention, and in fact, may even inhibit that retention.

We consider the following examples: (a) Merwin and Wickens (1991) compared data visualization with 2D and 3D (perspective) renderings. Subjects were asked questions that required the integration of information from a complex economic data set. Although the integrative questions were answered better with the more integrative 3D graph, a post test in which subjects were asked to describe the economic model that generated the data, revealed no difference between the two formats. Hence, the 3D perspective supported short term performance, but had no benefit for longer term retention.

(b) Aretz (1991) required subjects to perform a helicopter flight simulation in which they navigated through a virtual world of geometric objects. One group of subjects flew with a track-up (rotating) map. This "ego-referenced" feature alleviated the cognitive effort required to mentally rotate the map, and produced navigational performance that was superior to a group of subjects who flew with a fixed north-up map (world-referenced frame). Yet after the navigation was completed, when both groups were asked to draw a map of the environment through which they had traveled, the latter (fixed map) group showed better learning and retention of the position of the geographical features.

(c) Vicente (1990) compared two display designs for supporting diagnosis of complex system failures, a more standard design and an "ecological interface" that presented greater visualization of the relation between variables and their relation to underlying thermodynamic principles. While the ecological interface supported better diagnosis of system failures, it did not lead to better long term retention of specific values.

(d) Yallow (1980) has concluded that providing subjects with instructional material in the format (spatial-graphical vs. verbal-words) which is most compatible with their cognitive strengths, will lead to better

immediate comprehension of the material, but not to better long term retention.

(e) There has been some evidence from flight simulation research that simulator augmentation via guided training, that is more likely to produce error free performance (i.e., guiding the learner through the correct landing path), may produce poorer transfer or long term retention of the landing skills, once such augmentation is withdrawn (Lintern & Roscoe, 1980).

The previous five examples suggest that features designed to reduce the workload of processing and/or increase performance will not necessarily improve retention of the processed material. Two further lines of research bear on this dissociation between virtual reality and effort on the one hand, and learning on the other.

(f) Returning to the flight simulator environment, there is by now ample evidence that more realism is not necessarily more conducive to effective learning. For example, the presence of simulator motion, an enhanced source of sensory feedback, appears to help performance, but not transfer to the aircraft (Koonce, 1979; Lintern, 1987). Other research by Lintern and his colleagues, especially Lintern, Roscoe, and Sivier, 1990, suggest that added realism will not invariably improve learning. The reasons for these failures of realism probably lie in the diversion of the learner's cognitive resources to those elements of the simulation that are not essential to the fundamental understanding of flight dynamics, or in the case of motion, to information that may be of very little value. Similarly with VR, excessive realism and sensory experience, while "natural" can distract the learner from focusing attention on the key relations to be mastered.

Closed loop interaction and learning. The previous examples have all suggested that various elements present in virtual reality, while facilitating performance, may disrupt retention. However, there is one component, closed loop interaction, for which strong supportive evidence exists that it is learning enhancing. The irony here is that from the point of view of effort and performance, interactive participation generally produces short term costs, which are then compensated by later gains in knowledge. Here again, several examples can be cited. Kessel and Wickens (1982) found that in understanding tracking dynamics, subjects formed a better mental representation of those dynamics while actively tracking them (closed loop interaction with higher workload) than while passively monitoring an autopilot controlling the same system. The operational flying community has expressed a concern, with a similar interpretation, that higher levels of flight deck automation, designed to reduce workload and so render the pilot's interaction with the aircraft a cognitively simpler one, have left pilots out of touch with the momentary state of the aircraft and also have degraded the accuracy of their mental model of the flight dynamics (Wiener & Curry, 1980).

An experimental study by Von Wright (1957) supports this argument. Subjects tracked a maze with a series of dichotomous choice points. In the "active" condition, subjects could view ahead of each choice point, which leg was "open" and which was "closed." In the passive condition, subjects never made that choice but were forced to travel on the open leg. The results unambiguously indicated better retention for the active control group.

Collectively these diverse results may be summarized by a note of caution: various techniques have been advocated and successfully tested to reduce the cognitive effort of the user of computer interfaces; many of these techniques correspond to the features of virtual reality. While some of these might also improve long term retention of information that exists within the virtual world, some explicitly do not, and so the techniques of optimal

user interface design should not be applied wholesale to computer-based learning interface design.

ACKNOWLEDGMENTS

The author wishes to acknowledge Lt. Col. Dave Porter and Capt. Jay Horn for comments on this paper.

REFERENCES

- Aretz, A.J. (1991). The design of electronic map displays. Human Factors, 33(1), 85-101.
- Cohen, J.S. (1989). Visualization explores the depths of visual perception. Nat'l. Ctr. for Supercomputing Applications Newsletter, Access, Vol. 3(6), 11-12, University of Illinois at Urbana-Champaign.
- Harwood, K., & Wickens, C.D. (1991). Frames of reference for helicopter electronic maps: The relevance of spatial cognition and componential analysis. Int'l. J. of Avia. Psych., 1, 5-23
- Hutchins, E.L., Hollan, J.D., & Norman, D.A. (1985). Direct manipulation interfaces. Human-Computer Interaction, 1(4), 311-338.
- Kessel, C.J., & Wickens, C.D. (1982). The transfer of failure-detection skills between monitoring and controlling dynamic systems. Human Factors, 24, 49-60.
- Koonce, J.M. (1979). Predictive validity of flight simulators as a function of simulator motion. Human Factors, 21, 215-223.
- Lintern, G. (1987). Flight simulation motion systems revisited. Human Factors Society Bulletin, 30(12), 1-3.
- Lintern, G., Roscoe, S. N., & Sivier, J. (1990). Display principles, control dynamics, and environmental factors in pilot training and transfer. Human Factors, 32, 299-317.
- Lintern, G., & Roscoe, S.N. (1980). Visual cue augmentation in contact flight simulation. In S.N. Roscoe (Ed.), Aviation Psychology. Ames: Iowa State University Press.
- Merwin, D.H., & Wickens, C.D. (1991). 2D vs. 3D display for multidimensional data visualization: The relationship between task integrality and display proximity. Proc., 35th An. Mtg. of the Human Factors Society. Santa Monica, CA: Human Factors Society.
- Sweller, O., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load as a factor in the structuring of technical material. J. of Exptl. Psych.: Gen., 119, 176-192.
- Vicente, K.J. (1990). Supporting knowledge based behavior through ecological interface design. Univ. of Ill. at Urbana-Champaign unpublished doctoral dissertation. Department of Mechanical and Industrial Engineering, Urbana, IL.
- Von Wright, J.M. (1957). A note on the role of "guidance" in learning. Brit. J. Psychol., 48, 133-137.
- Wickens, C.D. (1991) Processing resources and attention. In D. Damos (Ed.), Multiple task performance. London: Taylor & Francis.
- Wiener, E.L., & Curry, R.E. (1980). Flight deck automation: Promises and problems. Ergonomics, 23, 995-1012.
- Yallow, E. (1980). Individual differences in learning from verbal and figural materials. (Aptitudes Research Project T.R. No. 13). Palo Alto, CA: Stanford, CA: Stanford University, School of Education.
- Zorpette, G. (1989). The main event. IEEE Spectrum, 26(1), 28.



Theories of Expertise as Models for Understanding Situation Awareness

Peter M. Crane, PhD
Aircrew Training Research Division
USAF Armstrong Laboratory

Abstract

Situation awareness is often understood as a skill unique to piloting aircraft and controlling other real time processes. Research intended to better understand or to measure situation awareness, however, has proven to be difficult. An alternative approach to investigating situation awareness is proposed using theories of expertise as models. These theories provide a mechanism which explain why pilots will maintain or fail to maintain situation awareness. Further, models based on theories of expertise provide testable hypotheses regarding quantifiable aspects of situation awareness and suggest training interventions to improve situation awareness.

Situation awareness is a term used by pilots and other operators of real-time processes to describe the operator's, "knowledge about his surroundings in light of his mission's goals," (Whitaker & Klein, 1988, p. 321). The concept is most often used to explain performance failures. A pilot who for no discernable reason does not respond to a threat such as an enemy aircraft or the ground, is said to have failed to maintain situation awareness. Cognitive psychologists have rapidly embraced the concept of situation awareness even though a workable definition is elusive (Fracker, 1988; Sarter & Woods, 1991).

Although the concept of situation awareness is undoubtedly useful, cognitive psychologists are experiencing great difficulty in defining the term and applying it to research. A major source of this difficulty is that situation awareness has no grounding in cognitive theory or previous research. Situation awareness is a pilot's concept used to explain why things go wrong. Applying empirical research tools to what is an intuitive concept has failed to shed much light on questions such as: what are the mechanisms for maintaining situation awareness, what are the variables which will allow one person to maintain situation awareness while another will not, is maintaining situation awareness a mental skill or ability by itself or is it the result of other cognitive skills? Attempts to understand situation awareness by directly studying pilots and other real-time process operators have thus far produced mixed results.

An alternative approach to understanding situation awareness is to review existing research and theory in cognitive psychology to identify phenomena which display the same characteristics as situation awareness. These phenomena can then be used as models to explore situation awareness and to understand the cognitive processes involved in maintaining or losing situation awareness. With this understanding, researchers will be better able to devise and validate training interventions to reduce the incidence of performance failures due to loss of situation awareness. One phenomenon which may serve as a model for understanding situation awareness is the study of expertise. Expertise can be defined as the ability to perform demanding tasks rapidly and error-free; compared to the expert, the novice's performance is

92-17676



slow, effortful, and prone to many errors (Anderson, 1983). Given that performance failures are often attributed to the loss of situation awareness, expert-level performance may be the equivalent to maintaining situation awareness for a highly demanding task. A review of the literature demonstrates that there are many similarities between situation awareness and expertise. Further, the extensive body of theory on expertise can shed significant light on the mechanisms of situation awareness and can provide many opportunities for research and development of training systems to improve situation awareness.

Comparison of Situation Awareness to Expertise

Endsley (1990) reported research on situation awareness in which pilots flew a series of simulated, tactical air-combat missions. In Endsley's studies, "At some random point in time the simulation is stopped and the cockpit and out-the-window displays are blanked. The pilot is asked a series of questions in order to determine his knowledge of the situation at that exact moment in time," (p. 42). Endsley found that ability to recall the location of other aircraft in the simulation was significantly related to mission success. Endsley's research on tactical pilots is highly reminiscent of a landmark series of studies on chess expertise conducted by Chase and Simon (1973). Chase and Simon found that the most successful players (experts) were able to recall the position of chess pieces on a briefly presented board while less successful players (novices) were not. The experts' ability to recall board position, however, was no greater than the novices' if the chess pieces were positioned at random. Chase and Simon assert that the experts' ability to rapidly encode and store positions resulted from their ability to recognize configurations and patterns among the pieces. Information available in the experts' long term memory allowed them to encode the chess board positions as a small set of interrelated configurations (i.e., "chunking") rather than as a large set of individual positions. Similar results have been found in comparisons of expert vs novice bridge players (Engle & Bukstel, 1978) and baseball fans (Chisi, Spilich, & Voss, 1979). Charness (1989) concludes from these results that changes in encoding speed and chunking efficiency support a hypothesis that increases in skill result from increases in the number of stored patterns in the expert's knowledge base and a significant increase in the expert's ability to access this knowledge. Feigenbaum (1989) goes further to summarize these and other findings as the knowledge principle:

A system exhibits intelligent understanding and action at a high level of competence primarily because of the specific knowledge it can bring to bear: the concepts, representations, facts, heuristics, models, and methods of its domain of endeavor. (p. 179).

Klein, Calderwood, and Clinton-Cirocco (1986) interviewed 26 midwest fire captains regarding the manner in which they establish situation awareness when arriving at a fire. All of the officers were highly experienced with a mean of 23.2 years as firefighters. Klein et al determined that decisions were made very quickly, less than one minute, in 78% of the cases studied. Most of the longer decision times were associated with a fire at a gas and oil storage complex. In this case, "the local firefighters who tried to contain the blazes could not be considered experts. They had no experience with a fire of this magnitude or with many of the specific problems encountered," (p. 557). Klein et al categorized each of 156 decisions as reflecting either a deliberated process in which the officer evaluated alternative approaches or

a prototype match process in which the officer assigned the fire to a specific category based on experience. Evidence of deliberative process was found for only 19% of the incidents and half of these came from the storage complex fire. The remaining decisions were made by rapidly matching the fire to a stored prototype. The fire captains then refined their mental model of the fire based on perceptual cues. In sum, Klein et al found that experienced fire captains under time pressure used their knowledge base of prototype fires to establish situation awareness.

Researchers studying situation awareness have stressed the need for pilots to rapidly encode and understand the significant events occurring in the environment (Sarter & Woods, 1991). Researchers studying expertise have found that these are the abilities which characterize expert level performance. Further, research has demonstrated that expertise is based on rapid access to an extensive knowledge base of domain specific information. The expert identifies the present situation as an example of a familiar prototype and selects a response accordingly. Fracker (1988) asserts that an identical mechanism is responsible for situation awareness: "Matched knowledge structures...provide the pilot's assessment of the situation and serve to guide his attention," (p. 102).

Skilled Memory as a Model for Situation Awareness

The theory of skilled memory has been proposed by Chase and Ericsson (1981) and Ericsson and Stazewski (1989) to explain the cognitive mechanisms responsible for expert level performance. These authors assert that expertise is domain specific, requires long practice, and frequently involves a mnemonic encoding strategy. From the study of experts in several domains, they postulate three principles of skilled memory.

1. "Experts use their knowledge structures in semantic memory to store information during skilled performance of some task," (Chase and Ericsson, 1981, p. 159).
2. "Expert memory involves organized and direct retrieval from long term memory," (p. 175). The two corollaries to this second principle are that experts know when to apply knowledge to a problem and that experts store intermediate knowledge states for future reference in directly accessible long term memory.
3. As memory skill increases, the time required to encode and retrieve task relevant information decreases (p. 184).

The theory of skilled memory asserts that experience in and of itself is not sufficient to produce expertise. Rather, the journeyman who is gaining experience is also learning to encode the significant elements of a given situation into a knowledge structure. The encoding process is self-addressable in that the associations used to encode a situation into a given knowledge structure also are the retrieval cues used to recall the information when a similar situation is encountered (Ericsson and Stazewski, 1989, p. 239). With experience, the journeyman learns to access the appropriate knowledge structure and abstract the relevant features. This process gets faster with experience until it appears to be automatic. Equating situation awareness with application of skilled memory provides several testable hypotheses about situation awareness and suggests opportunities for developing training interventions.

Implications of Skilled Memory Theory for Understanding Situation Awareness

1. Maintaining situation awareness is the product of expertise and not an independent skill or mental ability. Training designed to increase situation awareness directly will be fruitless. Training must be designed to increase the breadth and depth of the pilot's skill in tactical situations.

2. Situation awareness and expertise are specific to a knowledge domain. Pilots will have difficulty maintaining situation awareness in an unfamiliar tactical environment. Training which provides experience for unfamiliar events will be highly effective in improving a pilot's ability to maintain situation awareness. Houck, Thomas, and Bell (1989) conducted simulated, multiplayer air combat training exercises. Participants in these exercises were mission-ready F-15 pilots and air weapons controllers. For many tasks, participants rated the multiplayer, air-combat training as both valuable for increasing mission success and superior to the training received in the aircraft. Notably, these tasks were relatively unfamiliar because they are rarely practiced in the aircraft due to safety, security, and cost restrictions: tactics against four or more bogeys, chaff and flare employment, electronic countermeasures, all-aspect defense and, work with air weapons controllers. Tasks which are practiced in the aircraft were rated as not having significant training need: tactical formation, visual lookout and, basic fighter maneuvers. Based on skilled memory theory, it is hypothesized that a pilot's ability to recall events from a multiplayer air combat sortie will increase with increasing experience in similar engagements. Further, increasing ability to recall events will be correlated with increasing ability to maintain situation awareness and ultimately to increasing combat success.

3. The effectiveness of simulator-based training will be greatly increased by instructional support in assisting participants to understand, interpret, and encode the situations, actions, and results of simulated engagements. This assistance must be provided by more experienced pilots supported by well designed mission planning and debriefing tools.

The Aircrew Training Research Division of the USAF Armstrong Laboratory is currently developing the capability to support multiplayer, air combat simulation. Research will be conducted at Armstrong Laboratory to identify training strategies which will improve combat skills for tactical pilots including the ability to maintain situation awareness. Using existing cognitive theory and research as models for problems of interest to combat pilots will support the effective development of training procedures to improve combat readiness.

References

- Anderson, J. R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.
- Charness, N. (1989). Expertise in chess and bridge. In, D. Klahr & K. Kotovsky, (Eds.), Complex information processing: the impact of Herbert A. Simon, Hillsdale, NJ: Lawrence Erlbaum Associates.

- Chase, W. G. & Ericsson, K. A. (1981). Skilled memory. In, J. R. Anderson (Ed.), Cognitive skills and their acquisition, Hillsdale, NJ: Lawrence Erlbaum Associates.
- Chase, W. G. & Simon, H. A. (1973). The mind's eye in chess. In, W. G. Chase, (Ed.), Visual information processing. New York: Academic Press.
- Chiesi, H. L., Spilich, G. J., & Voss, J. F. (1979). Acquisition of domain-related information in relation to high and low domain knowledge. Journal of Verbal Learning and Verbal Behavior, 18, 257-273.
- Endsley, M. R. (1990). Predictive utility of an objective measure of situation awareness. Proceedings of the Human Factors Society 34th Annual Meeting, Santa Monica, CA: The Human Factors Society.
- Engle, R. W. & Bukstel, L. (1978). Memory processes among bridge players of differing expertise. American Journal of Psychology, 91, 673-689.
- Ericsson, K. A. & Stazewski, J. J. (1989). In, D. Klahr & K. Kotovsky, (Eds.), Complex information processing: the impact of Herbert A. Simon, Hillsdale, NJ: Lawrence Erlbaum Associates.
- Feigenbaum, E. A. (1989). What hath Simon wrought? In, D. Klahr & K. Kotovsky, (Eds.), Complex information processing: the impact of Herbert A. Simon, Hillsdale, NJ: Lawrence Erlbaum Associates.
- Fracker, M. L. (1988). A theory of situation assessment: Implications for measuring situation awareness. Proceedings of the Human Factors Society 32nd Annual Meeting, Santa Monica, CA: The Human Factors Society.
- Houck, M. R., Thomas, G. S., & Bell, H. H. (1989). The training potential of multiplayer air combat simulation. Proceedings of the Human Factors Society 33rd Annual Meeting, Santa Monica, CA: The Human Factors Society.
- Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid processing on the fire ground. Proceedings of the Human Factors Society 30th Annual Meeting, Santa Monica, CA: The Human Factors Society.
- Sarter, N. B. and Woods, D. L. (1991). Situation awareness: A critical but ill-defined phenomenon. International Journal of Aviation Psychology, 1, 45-57.
- Whitaker, L. A. & Klein, G. A. (1988). Situation awareness in the virtual world: Situation assessment report. Proceedings of the 11th Psychology in the Department of Defense Symposium, Colorado Springs, CO: U. S. Air Force Academy.



A Structural Measure of Learning for Ill-Structured Domains

John Landry, Ph.D.
Metropolitan State College of Denver

92-17677



Abstract

This paper proposes and illustrates a technique to assess structuring which is an essential characteristic of learning. Structuring is defined as the addition, association, retention, and application of new concepts or relationships in a schema. The proposed approach is useful for studies of managerial cognition since, unlike many techniques which require identification of a learner's schema concepts *a priori*, this measure can be applied to ill-structured managerial domains which do not provide the opportunity to derive concepts *a priori*. The metric is illustrated by developing cognitive maps of financial analysts' rationale for firm earnings. Analysts operate in an ill-structured environment and routinely produce documents which can be used to discover their assertions about the causes of firm earnings. An unobtrusive cognitive mapping method is applied to documents to discover and represent these assertions.

The contemporary military and civilian organizational environment is dynamic and increasingly interconnected. This circumstance challenges leaders to learn new behaviors in order to succeed and survive. However, the dynamic nature of organizational environments and the ill-structured decision-making tasks may obscure observation or measurement of learning since the state of theory development in command and managerial domains does not support use of measure requiring *a priori* discovery of concepts (Mintzberg, Raisinghani, & Theoret, 1976). This paper begins with the theoretical basis for the metric. Next, the structuring metric is developed. Finally, an example of its use will be presented.

Theory

Learning is defined as the process of acquiring new beliefs and applying them to guide action and to improve performance (Simon, 1983). The learning process may be represented as a general systems model consisting of input-process-output-feedback components. Inputs are cues from the environment; cognitive processes modify the learner's schema; outputs effect the environment through the learner's actions; and, feedback is applied to modify the schema if the outcome does not match expectations (Hastie, 1984). The model is simplified for this research by assuming a single learning cycle (Bateman & Zeithaml, 1989).

The central component of the learning model is the individual's schema. Schema are a network of beliefs consisting of nodes and links; nodes represent concepts and links represent a relationship between concepts. For example, Minsky (1981:98) proposed a "frame" representation of knowledge, following Bartlett's notion of schema; "We can think of a frame as a network of nodes and relations." Winston (1977:181), building on the notion that "frames" must represent facts, asserts that "a fact is a relationship together with a few things the relationship ties together in a meaningful way." Thus, the relevant unit of analysis is posited to be the relationship between concepts.

Clearly the formation of new concepts is important to learning; however, neither the origin of the concepts nor the novelty of the concept *per se* are included in the notion of structuring as developed in this paper. Anderson (1981:123) comments that "concepts are frequently taken as the unanalyzable building blocks or the primitives of cognitive representation." On the other hand, forgetting concepts might be plausible. Rumelhart, Lindsay, and Norman (1972:231) assert that "Excluding physical or physiological damage, we find no evidence whatsoever to support the notion that the material in LTM [Long Term Memory] can ever be deliberately forgotten." Thus, concepts,

once displayed in a domain, are posited to remain available indefinitely.

Learning produces several categories of schema change. For the purposes of this paper, schema development is divided into two categories following the dichotomous typologies of Carbonell, Michalski, and Mitchell (1983), Piaget (1960), and Simon (1983). First, structuring is defined as the addition, association, retention, and application of new concepts or relationships in a schema. Second, adaptation is defined as the transformation of existing knowledge in order to improve performance over repeated trials by instantiation of new data into extant categories or by the modification of extant schema relationships. Put another way, adaptation changes the functional relationship between concepts.

Many researchers consider structuring to be the most significant component of learning because it is the essence of problem solving wherein new approaches to a problem are developed (Simon, 1983; Norman, 1982; Michalski et al., 1983). Furthermore, structuring requires substantial cognitive effort, which implies a cost in either time or effectiveness (Norman, 1982; Simon, 1983). Structuring is also important since humans, unlike computers, do not have the capacity to directly copy and immediately apply knowledge (Simon, 1983). Thus, developing a greater understanding of structuring is significant because of its potential contribution to theory and practice.

A Structuring Metric

The notion of structuring is that a schema at time t_{n+1} is the sum of the schema at time period t_n plus the outcome of any learning process that occurred between time periods t_n and t_{n+1} . Simply stated structuring occurs when a relationship between entities is added or deleted; a relationship is a link between schema nodes. Thus, schema structuring, S , between two time periods, t_n and t_{n+1} , is given by:

$$S(t_n, t_{n+1}) = \text{Unique relations at } t_n + \text{Unique relations at } t_{n+1} \quad (1)$$

A unique relationship at t_n exists when the relationship is deleted between time periods t_n and t_{n+1} ; a unique relationship at t_{n+1} is the outcome of links being added between time periods t_n and t_{n+1} .

$S(t_n, t_{n+1})$ ranges from zero (0) to an upper bound equal to the maximum number of possible links. The number of links in a schema with N nodes is less than or equal to N -squared. However, by assumption, schema may not have loops (Axelrod, 1976). Thus, the maximum number of nodes adjusted for the potential of N nodes linking to themselves is $L_{\max} = N(N-1)$. Also, since managerial problems are ill-structured, the concepts included in a domain can change over time and, thus, the count of the number of nodes, N , is time-dependent. Following Rumelhart et al's (1972) notion of the persistence of memory, this paper posits that the appropriate metric for the total number of concepts in a domain, N_{domain} , is the union of all concepts observed in a domain. Substituting N_{domain} for N in the equation for $L_{\max}$ yields $L_{\max} = N_{\text{domain}}(N_{\text{domain}}-1)$. Dividing by two (2) gives the maximum number of undirected links between concepts.

Although $S(t_n, t_{n+1})$, N , N_{domain} , and $L_{\max}$ are absolute measures of structure, to facilitate comparison between different subjects' schema normalized structural measures are required because each subject may have different concepts in the same domain. Since data sources may include open-ended interviews and documents, one factor influencing the structuring measures is length of discourse; lengthy discourse is expected to possibly reveal a larger number of concepts and relations. Levi and Tetlock (1980) control for length by dividing the number of concepts by the logarithm of the number of words in the discourse sample. Levi and Tetlock (1980) report a significant ($r=0.57$, $p<.001$) correlation between the adjusted value and Schroder, Driver, and Streufert's (1967) cognitive complexity measure. Cognitive complexity is the number of dimensions or characteristics of a problem perceived by a decision maker. Equation 1 adjusted to account for discourse length is:

$$S_{adj} = S(t_n, t_{n+1}) / \log(\text{number of total discourse words}) \quad (2)$$

The denominator is the logarithm of the sum of the words in the discourse at time period t_n and t_{n+1} . Equation 2 is used in this paper as the measure of structuring.

An Example

The learning model introduced in the first section of this paper posits that structuring should be observed when the environment changes and/or when the performance does not match expectations. The following hypotheses are suggested:

- H(1): Environmental dynamism and outcome error are positively associated with cognitive map structuring.
- H(2): Environmental dynamism is positively associated with structuring.
- H(3): Outcome error is positively associated with structuring.

Subjects

This example is drawn from a larger research project (Landry, 1990) which examined the changes in financial analysts' cognitive maps as the result of forecast error and environmental dynamism. Financial analysts forecast firms' earnings performance; their data, analytical skill, tools, and techniques are equivalent to managers. Analyst's work product is a published forecast containing quantitative measures and text. Their quantitative forecasts are widely used in research supported by statistical methods. This research also uses a type of content analysis on their text to create a cognitive map of their rationale for earnings.

The sample consists of cognitive maps developed from publicly-available reports prepared 12 randomly selected analysts. A benefit of this approach is that documents provide an unobtrusive observation method in a natural organizational settings. The data consisted of two earnings forecasts which was prepared by each analyst at the beginning and end of the three-year observation period. Although the entire sample consisted of 24 reports, one pair of reports was identified as an outlier and is omitted from the remainder of this analysis.

Procedure

Axelrod's (1976) cognitive map technique operationalized the schema construct. Axelrod (1976:72) states that "A cognitive map is a certain way of representing a person's assertions about his beliefs with respect to some limited domain, such as a policy problem. The representation takes the form of a directed graph of points and the arrows between those points." Briefly, the cognitive mapping methodology consists of identifying the causal assertions which usually correspond to subject-verb-object. The subject and object indicate the nodes; a verb indicates a relationship.

Coding of the entire sample required over 200 hours to identify the 600 causal assertions and to reduce the data. The text portion of the twelve analysts' reports averaged 29.8 sentences per report and 25.0 causal assertions per report. The high percentage of causal assertions to sentences (i.e., 84 percent) suggests that analysts intend to communicate their reasoning for a firm's earnings. Two trained coders performed the causal analysis (interrater reliability = 0.74). To further facilitate the analysis, two coders (interrater reliability = .82) to reduce each subject and object to its morpheme.

The hypotheses were tested by operationalizing structuring, dynamism, and feedback. Structuring was computed by applying the S_{adj} metric developed above to each pair of analyst's cognitive maps. Multiple measures of dynamism and forecast error were used since no single measure is widely accepted. Following Tosi, Aldag, & Storey (1973), three measures of environmental dynamism were computed for the observation period. Sales dynamism and income dynamism were

computed the coefficient of variation of sales and income, respectively. The technology dynamism measure is an average ratio research and development expenses and capital expenditure to total assets. Performance feedback is represented by forecast error metrics (Armstrong, 1978). Following previous research on financial analysts' errors, three measures of forecast error were computed; mean absolute error (MA_Error), mean square error (MS_Error), and mean absolute percentage error (MAP_Error). Since the analysts forecasted earnings over different time periods, the measures were computed over the forecast period presented in their 1984 report.

Results

Table 1 reports results of tests of the research hypotheses. Prior to the analysis variables were examined and, when required, transformed to meet the assumptions of the hypotheses tests.

Table 1. Regression Analysis Results

Hypothesis and Independent Variable(s)	Overall F test	Prob. >	R-squared @ $p \leq 0.05$
H(1): Earnings_Dynamism & MA_Error	7.441	0.015	0.650
H(1): Earnings_Dynamism & MS_Error	7.067	0.017	0.639
H(1): Earnings_Dynamism & MAP_Error	8.431	0.011	0.678
H(1): Sales_Dynamism & MA_Error	0.431	0.664	NS
H(1): Sales_Dynamism & MS_Error	0.434	0.662	NS
H(1): Sales_Dynamism & MAP_Error	1.104	0.377	NS
H(1): Technology_Dynamism & MA_Error	1.718	0.240	NS
H(1): Technology_Dynamism & MS_Error	2.030	0.194	NS
H(1): Technology_Dynamism & MAP_Error	2.628	0.133	NS
H(2): Earnings_Dynamism	8.138	0.019	0.475
H(2): Sales_Dynamism	0.874	0.374	NS
H(2): Technology_Dynamism	2.660	0.137	NS
H(3): MA_Error	0.314	0.589	NS
H(3): MS_Error	0.309	0.592	NS
H(3): MAP_Error	1.177	0.306	NS

These results using models indicate some support for H(1) and H(2) at the $p \leq 0.05$ level. Since H(1) is simply the linear combination of the models posited by H(2) and H(3), the tests indicating support for H(1) may be confounded since the addition of another variable will usually increase the r-squared but the addition of a variable may not significantly contribute to the overall F (Tabachnick and Fidell, 1983). Tests indicated that the increased F of H(1) were not significant at the $p \leq 0.05$ level ($df=1,8$). Therefore, only the earnings-based model of H(2) is supported. The unstandardized coefficient for the earnings model is 9.10 ($t = 2.85$, $p \leq 0.0190$), and r-squared is 0.4748 (adjusted r-squared = 0.417). That is, when the objective environment dynamic, a statistically significant change in cognitive map structuring is found.

Discussion

The proposed measure of structuring relies on the notion that the relationships between concepts is central to understanding our world. These measures assume that a node cannot be linked to itself and that nodes, once observed, remain available although the relationships between nodes may change. Finally, to facilitate comparisons between subjects the measures are normalized for the length of discourse. The example illustrated environmental dynamism measure which supported hypotheses H(2) that posited a positive association between structuring and environmental dynamism. Moreover, the metric can be used with unobtrusive techniques like cognitive mapping to reveal the structure of cognition in normal, ill-structured organizational environments.

References

- Anderson, J. R. (1981). Concepts, Propositions, and Schemata: What are Cognitive Units. Nebraska Symposium on Motivation, 121-162.
- Armstrong, J. S. (1978). Long Range Forecasting: From Crystal Ball to Computer, John Wiley & Sons, New York, NY.
- Axelrod, R. (Ed.) (1976). Structure of Decision: The Cognitive Maps of Political Elites, Princeton University Press, Princeton, NJ.
- Bateman, T. S. & Zeithaml, C. P. (1989). The Psychological Context of Strategic Decisions: A Model and Convergent Experimental Findings. Strategic Management Journal, 10, 59-74.
- Carbonell, J. G., Michalski, R. S., & Mitchell, T. M. (1983). An Overview of Machine Learning. in Anderson, J. R., Michalski, R. S., and Carbonell, J. G. (Eds), Machine Learning, Morgan Kaufman, Los Altos, CA, 25-37.
- Hastie, R. (1984). Causes and Effects of Causal Attribution. Journal of Personality and Social Psychology, 46(1), 44-56.
- Landry, J. R. (1990). Environmental Dynamism, Forecast Performance, and Cognitive Map Changes: An "Optimistic" versus a "Pessimistic" Descriptive Model. Unpublished Doctoral Dissertation, University of Colorado.
- Levi, A. & Tetlock, P. E. (1980). A Cognitive Analysis of Japan's 1941 Decision for War. Journal of Conflict Resolution, 24(2), June, pp. 195 - 211.
- Minsky, M. (1981). A Framework for Representing Knowledge. in Haugeland, J. (Ed.), Mind Design, MIT Press, Cambridge, MA, 95-128.
- Mintzberg, H., Raisinghani, D. & Theoret, A. (1976). The Structure of "Un-structured" Decision Process. Administrative Science Quarterly, 21, June, 246-275.
- Norman, D. A. (1982). Learning and Memory, W. H. Freeman, New York, NY.
- Piaget, J. (1960). The Child's Conception of the World, Littlefield, Totowa, NJ.
- Rumelhart, D. E., Lindsay, P., Norman, D. A. (1972). A Process Model for Long-Term Memory. in Tulving, E. & Donaldson, W. (Eds.), Organization of Memory, Academic Press, New York, NY.
- Schroder, H. M., Driver, M. J., & Streufert, S. (1967). Human Information Processing, Holt, Rinehart and Winston, Inc., NY.
- Simon, H. A. (1983). Why Should Machines Learn. in Anderson, J. R., Michalski, R. S., & Carbonell, J. G. (Eds), Machine Learning, Morgan Kaufman, Los Altos, CA, 25-37.
- Tabachnick, B. G. & Fidell, L. S. (1983). Using Multivariate Statistics, Harper and Row, New York, NY.
- Tosi, H., Aldag, R., & Storey, R. (1973). On the Measurement of the Environment: An Assessment of the Lawrence and Lorsch Environmental Uncertainty Subscale. Administrative Science Quarterly, 27-36.
- Winston, P. H. (1977). Artificial Intelligence, Addison-Wesley, Reading, MA.



Performance decrements during sustained operations: The role of practice and task structure.

D.G. Humphrey¹, A.F. Kramer¹, R.R. Stanny² and K. Travis²¹University of Illinois at Urbana-Champaign ²Naval Aerospace Medical Research Laboratory

Abstract

The requirement for continuous performance for extended periods of time has prompted both laboratory and field studies of the effects of SUSOPS on the performance of a variety of perceptual-motor and cognitive tasks. In the present study we systematically examine the role that practice and task structure play in the moderation of SUSOPS effects in long term memory (LTM), working memory (WM), and visual search processes. Dependant measures include reaction time, accuracy and measures of event-related brain potentials (ERPs). The results suggest that perceptual processes are most susceptible to SUSOPS effects and that proper training may minimize these effects.

In recent years there has been a renewed interest in the effects of sustained operations (SUSOPS) on human performance. This interest can be attributed, in a large part, to the requirement for SUSOPS in both military and industrial settings. The requirement for continuous performance, often with little or no sleep, for extended periods of time has prompted both laboratory and field studies of the effects of SUSOPS on the performance of a variety of perceptual-motor and cognitive tasks (Hockey, 1986; Krueger, 1989). In the present study we systematically examine the role that practice and task structure play in the moderation of SUSOPS effects in long term memory (LTM), working memory (WM), and visual search processes.

A number of researchers have found impaired performance with SUSOPS, particularly when information was to be held in WM (Elkin & Murray, 1974; Polzella, 1975; Williams et al., 1966). However, other researchers have found either little or no decrement in WM processes with substantial sleep deprivation (Hamilton et al., 1972). Other studies have found decrements in LTM with extended periods of sleep deprivation, especially with tasks that had not been extensively practiced (Caille et al., 1972; Haslam & Abraham, 1987). The apparently inconsistent results for the WM tasks as well as the results obtained in the LTM tasks suggest that well practiced memory tasks may be somewhat resistant to SUSOPS effects (Fisk & Schneider, 1981).

In our study we investigated the role of practice and task structure during SUSOPS. In one version of a memory comparison task targets and distractors were consistently mapped (CM) throughout training. Extensive CM training leads to the development of automatic processing which is fast, often insensitive to capacity limits and not limited by WM (Shiffrin and Dumais, 1981). Given the lack of WM limits in automatic processing it is expected that SUSOPS will have little effect after extensive CM training. In another version of the memory search task the target and distractors changed roles over trials. This type of training is referred to as varied mapping (VM). Automatic processing does not develop during VM practice and therefore performance is slow and limited by WM. Thus, SUSOPS related performance decrements are expected in the VM version of the memory comparison task (Fisk & Schneider, 1981). Given that the development of automatic processing bestows a certain amount of resistance to SUSOPS on memory processes, it is of interest to determine the generalizability of this phenomenon. To that end we will also assess SUSOPS effects in well practiced CM and VM versions of a visual search task.

Methods

Subjects

Subjects were 12 dextral male Naval volunteers ranging in age from 21 to 26 years.

Memory search task

The goal of this task was to match a letter on the CRT with a set of previously memorized letters. At the beginning of a block of trials a set of letters was displayed and memorized. Subjects had a maximum of 10 seconds to memorize the set; pressing the space bar removed the letters if 10 seconds was not required. The removal of the memory set marked the beginning of a 30 trial block. Each trial consisted of one probe letter presented in the middle



of the screen. Subjects made one response if the probe matched a memory set item and another response if the probe did not match an item in the memory set. Memory search was carried out with memory sets of three and six members (SS3 and SS6 respectively) and under CM (e.g. targets and distractors did not exchange roles over training) and VM (e.g. targets and distractors switched roles across trials) training. Stimulus duration was 200 msec with a 1200 msec ITI.

Visual search task

The goal of this task was to compare a memorized target letter with a set of letters displayed on a CRT in order to determine if the target letter was present in the display. At the beginning of a block of 30 trials a single letter was presented, the target letter for the block. Each trial consisted of a set of letters (three or six) positioned evenly around an imaginary circle. The subject made one response if the target matched an item on the display and another response if there was no match. Visual search was carried out with display sets of either three or six members (SS3 and SS6 respectively) and under CM and VM training. Stimulus duration was 200 msec with a 1200 msec ITI.

Experimental apparatus and ERP recording

Subjects performed the tasks while seated in a sound and light attenuated booth equipped with an XT class computer with an EGA monitor. Responses were made using the "K" and the "L" keys of a standard keyboard. For training session six and beyond, 60 dB(A) white noise was binaurally presented over headphones. Electroencephalographic (EEG) activity was recorded from three midline sites (Fz, Cz and Pz) referenced to digitally linked and averaged mastoids. Vertical EOG was recorded from electrodes above and below the right eye. Horizontal EOG was recorded from electrodes lateral to each eye. EOG contamination of the EEG was compensated for off-line. All electrode impedances were kept below 10 kohms. The EEG and EOG were recorded with a 10 second time constant and filter pass-bands of 0.1 - 30 Hz. The data channels were digitized every 10 msec and were filtered off-line (-3 dB at 4.429 Hz, 0 dB at 14.29 Hz) prior to further analysis.

Procedure

Training. Subjects were run in pairs and performed two training sessions per day for Day 1 through Day 4, each session lasting approximately 1.5 hours. Each training session consisted of two iterations of each of the search tasks; an iteration consisted of 16 blocks of 30 trials. This yielded 960 search trials per task per session. Each iteration of a search task lasted 16 minutes.

SUSOPS. Following the training phase, subjects began the SUSOPS phase of the study. This phase began at 1800 of Day 4 and was completed by 830 the following day. Every 1.5 hours beginning at 1900 subjects performed a task session identical to the training sessions.

Dependant Measures. Reaction time (RT) and accuracy were recorded during both training and SUSOPS sessions with ERP recording being added in the SUSOPS sessions. The amplitude and latency of the P300 component was quantified on the subject's average ERP for each of the experimental conditions. The P300 latency was defined as the time at which P300 reached its maximum amplitude. P300 amplitude was defined as the mean voltage in a 50 msec window centered on the point of maximum amplitude. Analysis of the P300 component was performed at each of the electrode sites but the values reported here are for the Cz site.

We expect that the joint use of ERPs and performance measures will enable us to provide a fine-grained analysis of the changes in information processing that accompany the effects of SUSOPS (see Kramer & Spinks, 1991, for a review of ERP components). For example, since P300 latency is sensitive to stimulus evaluation processes but is relatively unaffected by response selection and execution processes, changes in P300 latency in response to SUSOPS in the absence of changes in RT would suggest that stimulus evaluation processes were affected. On the other hand, changes in RT in the absence of changes in P300 latency would localize SUSOPS effects to post-stimulus evaluation processes (e.g. response selection and execution processes). Finally, the amplitude of the P300 has been employed as a measure of attentional allocation. P300s have been found to increase in amplitude with increases in the amount of attention allocated to stimuli and tasks. Thus, changes in P300 amplitude as a function of variations in the level of SUSOPS will enable us to infer changes in attentional processes.

Results

In order to investigate the nature of SUSOPS's effect on automatic processing it is necessary to demonstrate that automaticity was attained during training in the CM conditions. After presenting the training performance data we will examine the effects of SUSOPS on performance and P300s. All of the data discussed are represented in Figure 1. All of the effects reported were significant with $p < .05$.

Training

Memory search. Main effects for RT included session, set size and mapping. Subjects responded more quickly with increased practice, were faster with small set sizes, and were faster in the CM than in the VM conditions. Importantly, a significant three-way interaction was obtained among session, set size and mapping factors for both RT and accuracy. The set size difference decreased to a greater degree in the CM than in the VM conditions with practice. In fact, by the end of session six the set size effect was three msec per item in the CM conditions and 38 msec per item in the VM conditions. The faster rate of learning and the smaller set size effect in the CM than in the VM conditions is evidence for the development of automaticity in the CM conditions.

Visual search. Main effects for RT included session, setsize and mapping. Subjects responded more quickly with practice. Subjects were also faster in the smaller display size conditions and in the CM conditions. The session by mapping by set size interaction was significant for RT. The RT difference between the two display sizes decreased to a greater degree for the CM than for the VM conditions with practice. The display size effect was 8 msec per item in the CM conditions and 15 msec per item in the VM conditions by session six. Thus, consistent with memory search training, the performance data indicated the development of automaticity in the CM but not the VM conditions.

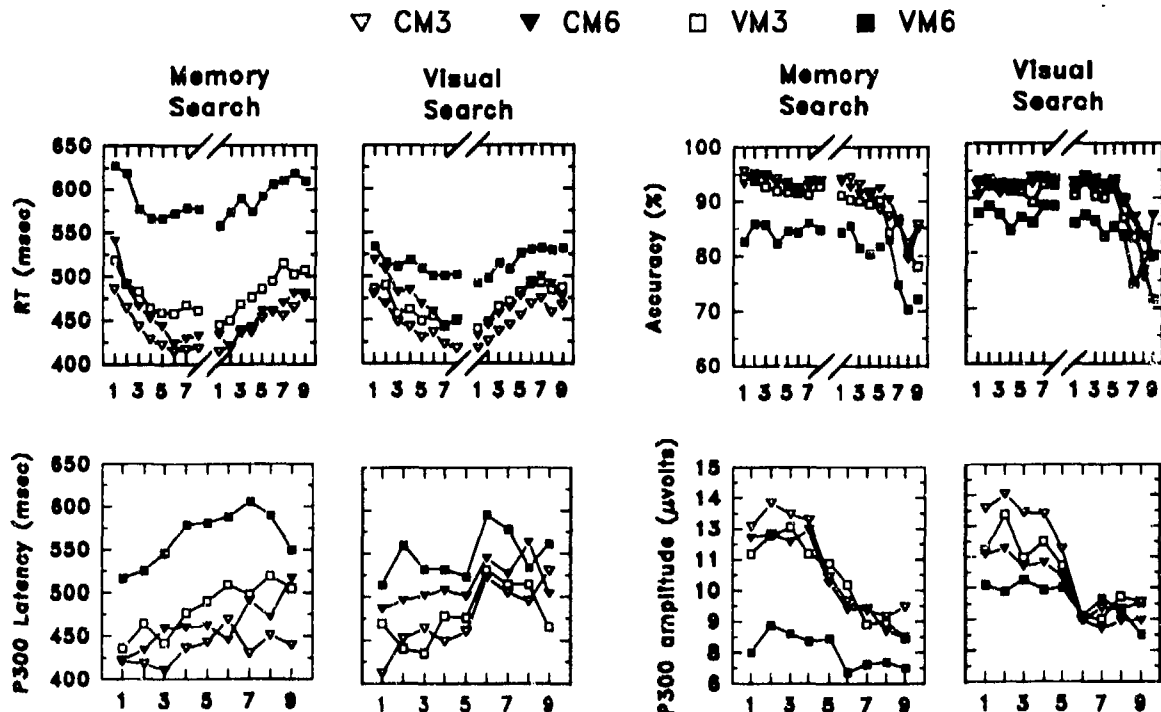


Figure 1: RT and accuracy for training (1 to 8) sessions. RT, accuracy, P300 latency and amplitude measures for SUSOPS (1 to 9) sessions.

SUSOPS

In the SUSOPS phase we examined the degree to which automatic processing which developed in the memory and visual search tasks was resistant to the detrimental effects usually associated with SUSOPS. We first report the

data for the performance measures. We then examine the effects of SUSOPS on P300 latency and amplitude.

Memory search performance. For the initial portions of the SUSOPS phase RTs were the same as those obtained during the final stage of training, but became slower with the increase in SUSOPS. The longer RTs with increased sleep deprivation occurred equivalently for each of the experimental conditions. However, although RTs increased with the number of SUSOPS sessions, the difference between set sizes was unaffected by sleep deprivation. The set size difference in the eighth SUSOPS session was five msec for CM conditions and 39 msec for VM. Thus, it would appear that both WM (VM conditions) and LTM (CM conditions) operations were immune to SUSOPS.

Visual search performance. Visual search results mirror those obtained in the memory search task. RTs increased with increases in SUSOPS. As with memory search performance the display set effect was uninfluenced by session. Set size effects were 11 msec for CM conditions and 15 msec for VM conditions in the eighth SUSOPS session. Thus, the visual search processes, whether automatic or non-automatic, were uninfluenced by increasing sleep deprivation.

To summarize, the performance results obtained in the visual and memory search paradigms suggest that WM, LTM, and visual comparison processes are uninfluenced by sleep deprivation, at least within the range of sleep loss investigated within the present study. On the other hand, there were substantial increases in RT in all of the experimental conditions in the visual and memory search tasks. Within the framework of serial stage models of information processing (Sternberg, 1966) such increases in intercept but not in search slopes would be attributed to either perceptual or response related decrements with increases in sleep deprivation. We now turn to an analysis of P300 data in an effort to further localize the processing deficits associated with SUSOPS.

Memory search ERPs. The P300 peak latency data shows a pattern similar to RT. The P300 latency differences between CM and VM conditions and SS3 and SS6 conditions are both significant as is the mapping by set size interaction. The lack of a session by mapping by set size interaction indicates that these differences remain throughout the SUSOPS phase. P300 amplitude decreased with increased sleep deprivation. P300 amplitude was larger for the CM conditions than the VM conditions and for SS3 than SS6 with increased SUSOPS.

The increasing latency and decreasing amplitude of the P300 component argues for a deficit in the processing that occurs prior to and including stimulus evaluation. The increase in P300 latency is comparable to the increases seen in the RT data. That these two temporal measures are related supports the conclusion that the effect of SUSOPS can be localized to perceptual processing. The decreasing P300 amplitude suggests a decreasing allocation of attention to the stimuli with increased sleep deprivation.

Visual search ERPs. The increase in P300 latency with increased SUSOPS resembles the increase in RT. The P300 component occurs earlier for CM than VM conditions and for SS3 vs. SS6 conditions. The mapping by set size interaction indicates that the display size effect is larger in the VM than the CM conditions. Consistent with RT, the display size by mapping interaction was not influenced by Session. There was a decrease in P300 amplitude with increasing SUSOPS. The CM conditions had larger P300 amplitudes than the VM conditions.

Discussion

Our original hypothesis was that the automatic processing that developed in the CM conditions during training would be less influenced by the increasing sleep deprivation in the SUSOPS phase than the non-automatic processing established in VM conditions. The results provide partial support for this hypothesis. The performance advantage that was exhibited for the CM conditions during training was maintained during the SUSOPS session. This advantage included both faster RTs and flatter search slopes for the CM than the VM conditions.

On the other hand, performance in both CM and VM conditions decreased at the same rate with progressive sleep deprivation. The joint examination of RT and P300 latency data suggests that this decrement in performance can be localized to perceptual rather than response processes. In addition, the decreased P300 amplitude with increases in SUSOPS sessions can be interpreted in terms of a reduced allocation of attention to process the probe stimuli with increased sleep deprivation. The decrease in P300 amplitude occurred at the same rate in CM and VM conditions.

These results along with the findings of insensitivity of visual and memory search slopes to progressive sleep

deprivation suggests that a) both automatic and non-automatic visual comparison and memory comparison and response related processes are somewhat resistant to the effects of SUSOPS, b) perceptual processes appear to be particularly sensitive to the detrimental effects of SUSOPS, and c) training that leads to automatic processing can be useful in minimizing SUSOPS effects - at least with respect to performance exhibited in tasks in which automatic processing does not occur.

Acknowledgement

This research was supported by the Office of Naval Technology. We gratefully acknowledge the support of Captain J. A. Brady, Commanding Officer, Naval Aerospace Medical Research Laboratory and his staff.

References

- Caille, E. J. P., Quidean, A. M. C., Girard, J. F. J., Gruber, J. C. & Monteh, A. C. (1972). Loss of sleep and combat efficiency: Effects of the work/rest cycle. IN W. P. Colquhoun (Ed.), Aspects of human efficiency: Diurnal rhythm and sleep loss. London: English Universities Press.
- Elkin, A. J. & Murray, D. J. (1974). The effects of sleep loss on short-term recognition memory. Canadian Journal of Psychology, 28(2), 192-198.
- Fisk, A. D. & Schneider, W. (1981). Control and automatic processing during tasks requiring sustained attention: A new approach to vigilance. Human Factors, 23, 737-750.
- Hamilton, P., Wilkinson, R. T. & Edwards, R. S. A. (1972). A study of four days of partial sleep deprivation. In W. P. Colquhoun (Ed.), Aspects of human efficiency: Diurnal rhythm and sleep loss. London: English Universities Press.
- Haslam, D. R. & Abraham, P. (1987). Sleep loss and military performance. In G. L. Belenky (Ed.), Contemporary studies in combat psychiatry. New York: Greenwood Press.
- Hockey, G. R. J. (1986). Changes in operator efficiency as a function of environmental stress, fatigue and circadian rhythms. In K. Boff, L. Kaufman and J. Thomas (Eds), Handbook of Perception and Performance, Vol. II. New York: Wiley and Sons.
- Kramer, A. F. & Spinks, J. (1991). Capacity views of human information processing. In J. R. Jennings and M. G. H. Coles (Eds), Psychophysiology of human information processing: An integration of central and autonomic nervous systems approaches. London: Wiley.
- Kreuger, G.P. (1989). Sustained work, fatigue, sleep loss and performance: a review of the issues. Work and Stress, 3, 129-141.
- Polzella, D. J. (1975). Effects of sleep deprivation on short-term recognition memory. Journal of Experimental Psychology: Human Learning and Memory, 204(2), 194-200.
- Schiffman, R. M. & Dumais, S. T. (1981). The development of automatism. In J. R. Anderson (Ed.), Cognitive Skills and their acquisition. Hillsdale, NJ: Erlbaum.
- Sternberg, S. (1966). High-speed scanning in memory. Science, 153, 652-654.
- Williams, H. L., Gieseking, C. G. and Lubin, A. (1966). Some effects of sleep loss on memory. Perceptual and Motor Skills, 23, 1287-1293.



Ergonomic Computer Work Environment

Senior Design Course

Cadet Brian Keller
 Cadet Todd McDowell
 Cadet Kelly Kirkpatrick
 Cadet Tom Becker
 Captain Jay Hori
 Captain Dean Orrell
 Major Mike Strevell
 USAF Academy

92-17679



Abstract

A class of Senior Air Force Academy Cadets designed an ergonomic computer workstation to complete their senior engineering course. The workstation prototype, called the 'Ergonomic Workstation Environment,' or EWE, was tested for comfort, mobility, and functional application by members of the Cadet Wing and Academy employees. The angles incorporated into the chair and monitor minimize stress on the body. Anthropometric data and lumbar support are incorporated in the design to provide greater comfort for the user. These design considerations will reduce the occurrence of fatigue and Repetitive Stress Injuries (RSI). The EWE fits through a standard door and is mounted on wheels for mobility. The chair is easily adjustable to fit users from the fifth percentile woman to the ninety-fifth percentile man. The EWE represents the computer work environment of the future.

Recent research has investigated the computer work area as an integrated system (Human Factors Society 1988). Many sources address the optimal design for consolidation of the monitor, chair, keyboard, and user (Grandjean 1987, Lueder 1986, Human Factors Society 1988, Corlett et al 1986, Kvalseth 1983 and Pheasant 1986). These studies emphasize an upright working position in relation to the monitor and seat. Attention to postural loading, optimum chair design, monitor tilt angle, monitor to user distance, and other ergonomic design factors were key components in these integrated designs. Data compiled from the above studies formed the research base for designing the 'Ergonomic Workstation Environment' or EWE.

The EWE design demonstrates integration of ergonomic principles and anthropometric data, and serves as a prototype for future research. The purpose of the EWE is to offer office and

industrial computer workers an ergonomically enhanced alternative to the traditional desk and chair computer environment.

The EWE was built using anthropometric research from a National Aeronautics and Space Administration (NASA) study which focused on officers and officer candidates between the ages of 23 and 35 from 1950 to 1973. In accordance with the NASA study, the EWE design adjusts for anthropometric measurements from the 5th percentile woman through the 95th percentile man, conforming to military design standards.

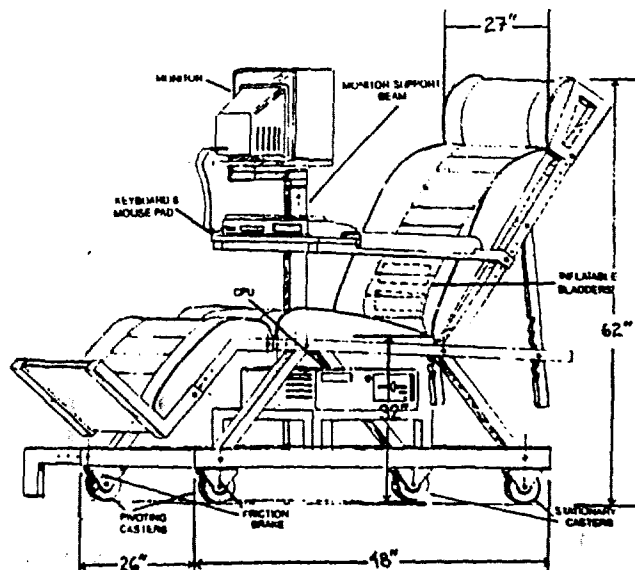
Overall Design

The EWE focuses on integrating the seating area, monitor, keyboard, armrests, and central processing unit into an ergonomically designed system. The design frees the user from the constraints of a desk, and instead cradles him or her in a unit designed for health and comfort. Once adjusted, the EWE would work well for a handicapped person who's movements are limited.

Our goal in the monitor arm design was to insure a correct viewing angle and viewing distance regardless of user size. The arm can swing 180 degrees to adjust the viewing distance from 12 to 43 inches, and can easily support a 40 pound monitor.

When adjusted in place, The keyboard and mouse platform is an extension of the armrests allowing wrist support while typing, and reducing the possibility of Carpel Tunnel Syndrome. The left armrest is adjustable up and down to allow easy access to the EWE. The CPU rests on a platform underneath the seat. Both the power bar and the disk drive are accessible to the user from a seated position. The EWE fits through standard doorways, and is compatible with many different types of personal computers.

Figure 1. The EWE: Specific Components and Overall Size (USAF Directorate of Visual Information Services, Graphics Division).



The EWE can be adjusted in several areas (see Figure 1): (1) the seat pan tilts front to rear; (2) armrests adjust 7-12 inches vertically, and extend 1-16 inches forward; (3) knee to footrest length extends from 17.75 to 25.25 inches; (4) thigh to knee length extends; (5) the headrest adjusts vertically; and (6) the chair compensates for any instability with outriggers that extend a maximum of 15 inches. The headrest tilts forward as a final adjustment parameter. The EWE also includes inflatable air bladders in the lumbar region for additional back support. From the horizontal, the seat back and seat angles adjust 95-125 degrees and 0-25 degrees respectively. The head rest tilts forward a maximum of 45 degrees from the seat back angle.

Derived mostly from Grandjean and Lueder studies, the adjustment parameters in the preceding paragraph represent the final design parameters for the EWE. These numbers are a compromise between the "zero gravity position," which describes the body's natural posture in space, and the upright computer stations mentioned earlier.

The EWE is reliable because of its simplicity. The air bladders are inflated manually, and there are no electronic parts to malfunction. The EWE frame is constructed of square aluminum tubing and upholstered with a breathable and durable fabric.

In summary, we changed the old desk and chair combination to an integrated system of comfort and mobility. Because the workstation is ergonomically designed, long term stress injuries will be reduced.

Method

Subjects

The ideal user for this work station is someone who does nothing but interface with their computer. Such jobs would include telephone receptionists, telecommunication agents, or travel agents who work using telephone headsets.

The test subjects were secretaries and instructors from five different academic departments at the Air Force Academy who used the EWE during a one to two day period. Each test subject gave feedback by either questionnaire or interview. Since secretaries in the Academy departments had to interact with the whole office and perform tasks not requiring a computer, we were limited in our ability to obtain ideal test results.

Apparatus

The test materials included the EWE, word processing software, and the evaluation questionnaire designed by the Engineering 410 students in conjunction with the USAF Academy Behavioral Science Department.

Procedure

The criteria used to analyze the performance of the EWE included ease of adjustability, comfort, and utility. These areas were addressed through the questionnaires completed by each subject after their test period with the EWE. Feedback included both qualitative and quantitative data.

Results

The qualitative results of our testing were recommendations for improvements in design which are explained in the discussion section. The quantitative results were divided into two main areas: adjustability and comfort. The adjustability scale ranged from 1 being very difficult to adjust to 5 being very easy to adjust, whereas the comfort scale ranged from 1 being extremely uncomfortable to 7 being extremely comfortable. The results are shown in Table 1 and Table 2.

Table 1

Mean Adjustability Scores from The Test Subject Questionnaire

EWE Component	Mean Score
Seat Back	4.75
Monitor Arm	1.75
Monitor Tilt	4.25
Footrest	3.25

Table 2

Mean Comfort Scores from The Test Subject Questionnaire

Part of the Body	Mean Score
Back	5.20
Neck	4.20
Wrists	4.40
Eyes	5.40
Shoulders	5.60
Legs	5.40

Discussion

The height of the EWE from the ground, 32", makes it difficult to enter and exit. Addition of a step welded to the frame would be the simplest solution to this problem. An additional solution would have the EWE redesigned closer to the ground. An improved monitor arm should be spring loaded to allow easy, accurate adjustment with one hand. The arm should be mounted further away from the right armrest and keyboard to allow greater adjustment of both. A safety strap for the monitor would prevent potential injury or damage should the monitor platform tilt too far forwards or backwards.

Improvements suggested by test subjects included reducing the EWE's overall size, and placing the CPU in a position where the disk drive is more easily accessible. Adjustability of the distance between the armrests would allow the user to adjust for comfortable elbow separation. We recognize other flaws may exist in the prototype design and trust future designers will find our research and the EWE beneficial in the continued development of ergonomic computer environments.

References

- Corlett, Niger, John Wilson, and Ilija Manenica (1986). The ergonomics of working postures: Modes, methods, and cases. London: Taylor and Francis.
- The flogiston chair: an environment for letting go. (1989). Austin, TX: Flogiston Corp.
- Grandjean, Etienne. (1987). Ergonomics in computerized offices. London: Taylor & Francis.
- Human Factors Society. (1988). Optimum work station angles. American national standard for human factors engineering of visual display terminal workstations (ANSI/HFS 100-1988). Santa Monica, CA: HFS.
- Kvalseth, Tarald O. (1986). Ergonomic of workstation design. Boston, Butterworth.
- Lueder, Rani. (1986). The ergonomics payoff: Designing the electronic office. New York: Nichols.
- MIL-STD-1472C. (1981, 2 May). Human Factors Engineering for Military Systems.
- NASA. (1978). Anthropometric source book: Volume I: Anthropometry for designers.
- Pheasant, Stephen. (1986). Bodyspace: Anthropometry, ergonomics, and design. London: Taylor and Francis.



Development of an Instructional Tutor
for the T-41 Flight Program

Cadet Steven A. Peeples
Captain Jay A. Horn
USAF Academy
Captain William A. Jones Jr.
Brooks AFB, Texas

Abstract

The purpose of this study is to use the Microcomputer Intelligence for Technical Training (MITT) software to develop an instructional tutor for the T-41 flight program at the United States Air Force Academy. The tutor was subsequently tested to determine if it could teach a novice the proper knowledge structure to become an expert in the domain. A modification was made on the Spilich et al. (1979) study of the effect of domain knowledge on recall to test the effectiveness of the tutor. Currently the tutor has not been completed, however, it is the authors' opinion that the MITT tutor will provide the same knowledge structure needed to link the causal events together making them more meaningful. This will allow the MITT subjects to recall domain related information from a passage better than those who have not used the tutor.

At the United States Air Force Academy (USAF) there is a statue inscribed with the words "Man's flight through life is sustained by the power of his knowledge." This quote reflects the importance of a good education for professional officers as well as the general society. The Air Force has a vested interest in the education level of its members, and more importantly the application of this knowledge.

This study will explore imparting the maximum amount of instruction to the individual in the minimum amount of time using an intelligent tutoring system (ITS). The Microcomputer Intelligence for Technical Training (MITT) software was used to develop the T-41 trouble shooting domain and gather data concerning its implementation. MITT is designed to teach a domain developed through the use of the MITT Writer program (MWRITER). These software packages were developed by Galaxy Scientific Corporation for the U.S. Air Force's Armstrong Laboratory Human Resources Directorate (ALHR) at Brooks Air Force Base in San Antonio, Texas.

To test this system we decided to apply the findings of high and low domain knowledge level as they pertain to recall. Spilich, Versonder, Chiesi, and Voss (1979) found individuals with high domain knowledge recall more domain related information from a scenario than a person with comparatively low domain knowledge. This research used a baseball domain, and in addition

92-17680



to recalling more of the domain related information high knowledge subjects recalled more goal oriented information.

Several other studies support these findings. Chiesi, Spilich, and Voss (1979) found that people with a higher level of domain knowledge use this higher knowledge to process, store, and retrieve the information. They state high knowledge groups use a working memory system allowing the individuals to keep track of the higher level goal orientated states of the domain.

Schneider, Koerhel, and Weinert (1989) tested children to determine if aptitude or domain knowledge level in soccer had a greater influence on recall of material. Their results state that "domain specific knowledge affected text recall." They further concluded that prior knowledge level had more of an effect on the subject's performance than their "general intellectual ability." This allowed the subjects to use "domain relevant patterns automatically" to work on the task they were presented with.

We feel this research serves as a good base to measure the MITT tutor. We plan to increase the domain knowledge structure, with the MITT tutor, of those individuals with low domain knowledge to be comparable to those with a previously high domain knowledge.

Preliminary Flight Training Program

The preliminary flight training program at the Air Force Academy is designed to provide cadets with a base of flight fundamentals to help them graduate from Undergraduate Pilot Training (UPT). This course consists of fifteen flights supplemented by classroom instruction on regulations and publications. This course is designed to teach the student how to control a Cessna T-41 aircraft in solo flight. The student must obtain a high level of proficiency to have the skills necessary to pilot the aircraft safely. Graduates of this program are then qualified to enter UPT.

We have chosen the T-41 program as the domain the tutor will instruct. We will use MITT to teach students procedure based trouble shooting of the T-41 aircraft. Through this research we hope to develop a worthwhile supplement to the T-41 program which will enable the students to learn more quickly, and with a better rate of recall. Additionally, we hope to make some general inferences about the utility of computer based instruction.

Statement of the Research Questions

Given that the United States Air Force Academy desires to produce officers who perform well in Undergraduate Pilot Training (UPT), would the use of an ITS during training help accomplish this goal? If an ITS can help produce cadets with higher aptitudes for flying, what are the most effective ways to use them? Additionally, can the increase in aptitude of the flying skills used in the T-41 program translate to UPT? Finally, can the Air Force produce more effective, more efficient, safer, and happier pilots using ITSs?

Statement of the Hypotheses

Preliminary research has shown that a student teacher ratio of 1:1 increases student's learning. Bloom (1984) found that students taught material under the 1:1 ratio scored in the 95th percentile of students taught in a group. We expect students trained using ITSS, and in particular MITT, will have higher scores on a post training test, and will recall this information longer. This can correlate to student pilots (on average) taking less time to reach the proper proficiency level. Students should also report higher levels of satisfaction with the course since they can progress at their own speed. Ultimately pilots trained and kept current using ITSS will be better pilots as witnessed by their elevated safety record, self report of satisfaction, and combat performance.

Method

Subjects

The subjects involved in this experiment consisted of cadets at the USAF Academy. Half of the participants were required to have a high domain knowledge, completion of the AV-441 Flight Indoctrination Course, and the remainder to have low domain knowledge, limited flight experience including military incentive rides and orientation courses.

Since cadets are under a high level of stress, they make good subjects for this study, as this stress compares to the stress experienced in UPT. Subjects volunteered to participate in this experiment, and ranged in age from 20 to 23.

To limit variability and increase the effort put into the experiment only pilot qualified cadets were asked to participate. The time given to this experiment is willingly spent by those trying to complete flight programs as well as those that have already completed the course. The subjects were told if they wished they could get a copy of the MITT program, for personal use to keep abreast of their flying skills.

Apparatus

For half of the subjects of both high and low domain knowledge groups, MITT was employed to give the subjects an increase in domain knowledge. The program was run on a Unisys 386 personal computer with a EGA, color graphics monitor.

Data was gathered using paper and pencil tests, a subjective evaluation, and an oral simulated problem task. The first test was given to assess the subject's knowledge before the training began. This test measures basic knowledge of the T-41 domain and allows verification of the subject's domain level group assignment. This test was given again to those completing the MITT training to assess the information learned. The oral scenario was then given to assess the recall of information in a fault diagnosis task as the dependant variable.

After the oral test was given a subjective questionnaire was given to those who had used the MITT training. This questionnaire was given to evaluate the subject's feelings about the training to assess the goal of providing enjoyable training. Finally, the subjective questionnaire was given again a week after the oral test had been administered. This test was used to evaluate the long term recall of the domain related information from the scenario of the MITT users against the subjects not using MITT.

Procedure

These tests were developed to assess the ability of MITT to provide fast, efficient, and interactive fault diagnosis training. The experiment was designed to minimize the time necessary for completion since cadet time is often limited.

As the participants were solicited to be subjects they were given the pre-test to determine domain knowledge. The subjects were placed into one of four groups. The first determinate of the proper group was T-41 experience and the preliminary test score. After the subject was determined to have high or low domain knowledge they were randomly assigned to the either the MITT group or the Non-MITT group. These groups are shown in Table 1. After the subjects were assigned to the respective groups the oral testing began.

Table 1
Detail of Subject Breakdown

		Training	
		MITT	NONE
Domain Knowledge	High	15	15
	Low	15	15

Note: Numbers represent the number of subjects in data subset.

The first subjects to be tested were those that did not required MITT training. These subjects were tested in the following manner. One subject at a time was brought into a dormitory room serving as a testing facility for all of the subjects. The subjects were read a scenario containing a trouble shooting task. The subject was then given a subjective questionnaire consisting of 25 fill in questions.

The subjects requiring MITT training were given up to four training periods, the combination of all lasting no longer that one hour. The total training time length was determined by how quickly the subjects completed the four assigned training blocks. Each training block was designed with about 10 minutes of material. After the subjects completed their MITT training they underwent the same testing procedures as outlined above.

In addition, the subjects given the MITT training were given the subjective questionnaire upon completion of the oral test. The questionnaire has seven questions relating to the training

which are evaluated on a five point Likert type scale. The questions range from the subject's evaluation of the effectiveness of the training provided to MITT's use of the graphic interactions. All of the subjects were given the questionnaire again one week after they initially were read the trouble shooting scenario to evaluate recall abilities.

Results

Currently the project is being held up by the processing of information into the tutor shell. Four scenarios have been identified by an expert, a T-41 instructor pilot, as those that represent a good cross section of the T-41 systems.

These four situations will be put into the tutor shell and subjects will be run in early February. However, several problems have been identified as being inherent to the T-41 tutor. The first problem that has been identified is that pilots in the Air Force are taught to leave the trouble shooting to maintenance. The pilots are instructed to isolate the problem and land as soon as possible. While this ability to just terminate the mission is not as readily accomplished during a wartime mission the T-41 is not a complicated enough system to lend itself to trouble shooting. This is the second problem in the tutor. The T-41 has a training mission and students are not encouraged to spend time analyzing the situation instead of simply landing the plane.

Additionally, the Air Force has determined to change the preliminary flight training aircraft. Currently the Air Force is in the process of acquiring a new aircraft to replace the aged Cessna T-41. This decision will render the MITT T-41 tutor useless, however, further tutors could be developed to maximize training and minimize cost. Subsequently, the lessons learned from the development of this tutor could help shape tutors for the new system.

References

- Bloom, B.S. The two sigma problem: the search for methods of group instruction as effective as one to one tutoring. Educational Researcher. vol 13 (6). 1984. pp. 4-60.
- Chiesi, Harry L., George J. Spilich, and James F. Voss. Acquisition of domain related information in relation to high and low domain knowledge. Journal of Verbal Learning and Verbal Behavior. 18. 1979. pp. 257-273.
- Schnieder, Wolfgang, Joachim Koerhel, and Franz E. Weinert. Domain-specific knowledge and memory performance: A comparison of high- and low- aptitude children. Journal of Educational Psychology. vol. 81 no. 3. 1989. p. 306-312.
- Spilich, George J., Gregg T. Vesonder, Harry L. Chiesi, and James F. Voss. Text processing of domain-related information for individuals with high and low domain knowledge. Journal of Verbal Learning and Verbal Behavior. 18. 1979. pp. 275-290.



Improving Visual Acuity Performance

Jennifer L. Hughes
Leray L. Leber, PhD
Amy L. Walker, O.D.

United States Air Force Academy

92-17681



Abstract

There is much evidence to suggest that unaided visual acuity is not an unalterable physiological process. Psychological processes allow some degree of pupil control, blur interpretation, accommodative flexibility, and other cognitive mediation of acuity. This research was designed to quantify visual acuity enhancement through a pre-test/post-test format. Seven Air Force Academy cadets were trained using the Acuity Performance Training software. All seven subjects manifested improvement in their vision, six in both unaided acuity and augmented recognition with less refractive correction. Training with this software might be used by any cadet to improve visual acuity performance prior to commissioning physical examinations.

Visual acuity is a primary determinant of whether a potential aviator is allowed to enter pilot training or maintain pilot qualification. A primary source of United States Air Force pilot candidates is the Air Force Academy and, accordingly, one of its goals is to both recruit potential pilots and have graduates enter pilot training. This requires that Academy cadets have good eyesight upon entrance and that they maintain their acuity during their four-year enrollment. Eighty-five percent of the 945-member Class of 1989 entered with pilot qualifying 20/20 visual acuity, however, only 59 percent demonstrated that acuity in their commissioning physicals.

Myopia, commonly referred to as "nearsightedness", is a visual anomaly in which parallel rays of light entering the eye, with accommodation relaxed, focus in front of the retina. In the majority of cases, myopes function normally with some degree of blurred distant vision or is corrected through the use of lenses that diverge parallel light rays to compensate for the eyes' excessive optical refraction.

The degree of refraction or accommodation required for clear distant vision is measured in diopters (D), the reciprocal of the corrective lens' focal length in meters. A 3.0D myope cannot focus beyond a distance $1/3\text{m}$ or 33cm in front of the eye, a distance called the far point. A myope without correction cannot attain a conjugate retinal image of targets beyond the far point. Refractive error is measured in one of two ways. One method requires the observer to attempt to recognize letters through progressively changing lenses until perfect performance is achieved. The second and more precise method requires the observer to attempt letter recognition through the same

progressively changing lenses while the eye's ciliary muscles are temporarily numbed (cycloplegia).

A common test of visual acuity and myopia is to have an examinee attempt to recognize individual letters on a Snellen Chart containing rows of high-contrast solid-black alphabet letters printed on a white background. The letters vary in size and are typically viewed from 20 feet (6.1 m). The examinee's resolving abilities are confirmed once they discern a predetermined percentage of letters of one size.

Acuity is often expressed as a ratio of a standard distance (20 feet) to the distance of lines of characters that subtend 1 minute of visual arc. Normal acuity, expressed as such a Snellen fraction, is 20/20. An acuity of 20/40 means that for the examinee to recognize a line of characters at 20 feet (6.1 m) the stroke width of the letters must subtend 2 minutes of visual arc such that an emmetrope with no visual anomalies could identify them at 40 feet (12.2 m). An acuity of 20/10 is the lower human limit, and a person with 20/400 acuity is defined as legally blind (Haber & Hershenson, 1980).

One study suggests that heredity might not be a primary determinant of myopia (Young et al., 1969). This study found the major difference between the older and younger subjects was reading and other near work, that this was the likely cause for the development of myopia in children. Additionally, several studies have accepted that effective vision is a function of learned visual and perceptual habits in a healthy organism as well as the optical characteristics of the eye, thus taking a behavioral approach to improving visual acuity (Ricci & Collins, 1981; Birnbaum, 1981).

Several optometrists (Lanyon & Giddings, 1974) have observed patients who perform far better in the real world than their clinically assessed acuity would allow. Bates (1920) observed improvements in the visual acuity of patients after training. Lastly, the conditionability of physiological functions has been established (ie- blood pressure, heart rate, body temperature; Bandura, 1969). Thus, clinical-behavioral precedent exists for exploring the role that conditioning plays in the modification of myopic acuity decrement.

Despite the disagreement as to its usefulness or rendering mechanisms, vision training has been used for decades. As early as World War II, combat pilots received visual training in rapid recognition with tachistoscopically flashed profiles and silhouettes of enemy and allied aircraft (Renshaw, 1945). Provine and Enoch (1975) trained subjects to use internal performance criteria independent of visual feedback to initiate and maintain accommodative responses. Several investigators have demonstrated the presence of voluntary accommodation in a variety of tasks (Cornsweet & Crane, 1970, 1973; Trachtman, 1978). Giddings and Lanyon (1974) found an average of 0.25 D less myopia in subjects who participated in visual acuity training.

Feedback obviously performs an important role in learning accommodative control. The goal of the feedback in the case of vision training is to place the mechanisms of perception under voluntary control. Further studies support these findings

(Giddings & Lanyon, 1974; Collins et al, 1981). Moreover, specific findings regarding the most effective method of training suggests that making the stimuli progressively smaller (called fading) is the most important part of this training technique although contingent feedback and fading may be slightly more effective than fading alone (Collins, Epstein, and Hannay, 1981).

Both extravagant claims and forthright condemnation of visual acuity training have been based on an incomplete understanding of the physiological and psychological mechanisms involved in acuity enhancement. Although there is little agreement regarding the etiology and the exact basis of myopia, improved visual acuity has been achieved using behavioral training techniques. Whether it is physiological changes, modified processing, motivation, or simply practice, behavioral training seems to improve visual acuity in some myopes.

Method

Subjects

Seven male cadets ranging in age from 18 to 22, Academy freshmen through seniors, served as subjects. The members of this group were referred by the Cadet Optometry Clinic based upon the results of their tested monocular visual acuities and refractive error.

Apparatus

An IBM-compatible computer was used to run the copyrighted Acuity Training Program software. This software randomly generated letters which formed a vision chart line. The display ranged from 20/12 to 20/120 when viewed from the distance of 20 feet (6.1 m). Included in this software was the capability to select the specific letters that were more difficult (ie- O, C, D) in order to practice solely with them. Subjects responded to each series of letters by typing keyboard inputs. If any responses were incorrect, a beep sounded and the correct letter was displayed 33 percent larger higher on the screen. When all the responses were correct, the next board appeared one increment smaller (20/38 from 20/39); if one response was incorrect, the next board appeared at the same size; and if more than one response was incorrect, the next board appeared one increment larger (20/40 from 20/39).

Procedure

The seven subjects were each trained for five-consecutive fifty-minute training sessions. The subjects sat 20 or 10 feet away from the screen depending upon their acuity. The initial training was conducted at the acuity rating provided by the preceding Cadet Optometry Clinic examination. The degree of myopia was quantified both by noting the smallest line of Snellen letters that could be monocularly recognized with no augmentation

as well as recording the minimum diopters of lens refraction correction required for the subject to recognize 20/20 letters. All subsequent training sessions were conducted at the final letter size from the previous session. At the end of the five sessions, the subjects were retested at the Cadet Optometry Clinic to determine whether an improvement had occurred in either monocular acuity or refractive error.

Results

In comparing the pre- and post-test results, all seven subjects experienced measured acuity improvement in at least one eye, three in both. Additionally, six of the seven subjects demonstrated refractive error improvement of at least 0.25 D in at least one eye. Specific results are annotated below:

<u>Subject</u>	<u>Eye</u>	<u>UNAIDED ACUITY</u>		<u>REFRACTIVE CORRECTION REQUIRED FOR 20/20 RECOGNITION</u>	
		<u>PreTng</u>	<u>PostTng</u>	<u>PreTng</u>	<u>PostTrg</u>
1	Right	20/25	20/20*	-0.25	-0.25
	Left	20/20	20/20	-0.50	-0.25*
2	Right	20/25	20/20*	-0.50	-0.50
	Left	20/30	20/25*	-0.75	-0.50*
3	Right	20/40	20/30*	-0.50	-0.50
	Left	20/50	20/50	-1.00	-0.75*
4	Right	20/70	20/70	-1.50	-1.25*
	Left	20/100	20/70*	-1.50	-1.25*
5	Right	20/80	20/80	-1.50	-1.50
	Left	20/50	20/30*	-0.75	-0.75
6	Right	20/100	20/80*	-1.50	-1.25*
	Left	20/100	20/80*	-1.50	-1.25*
7	Right	20/250	20/200*	-3.25	-3.25
	Left	20/250	20/200*	-3.50	-3.25*

* denotes training enhancement

Discussion

The results of this study could have far reaching affects as they relate to the United States Air Force. Indeed, since part of the flight qualification criteria includes visual acuity, monocular acuity as well as refractive error, the success achieved through this type of vision training might open up aircrew training to a whole new population. The Air Force Academy Association of Graduates has expressed an interest in providing funding to make this software and training instruction available to every interested USAFA cadet. As a result of this training, the Air Force might have a larger pool of candidates from which to select their future aircrew members. This could only lead to a more skilled and better prepared Air Force in the years to come.

References

- Bandura, A. (1969). Principles of Behavior Modification. New York: Holt, Rinehart and Winston.
- Bates, W.H. (1920). The Bates Method for Better Eyesight Without Glasses. New York: Holt.
- Birnbaum, M.H. (1981). Clinical management of myopia. American Journal of Optometry and Visual Development, 44, 554-559.
- Collins, F.L., Jr., Epstein, L.H., & Hannay, H.J. (1981). A component analysis of an operant training program for improving visual acuity in myopic students. Behavior Therapy, 12, 692-701.
- Cornsweet, T.N., & Crane, H.D. (1970). Servo-controlled infrared optometer. Journal of the Optical Society of America, 60, 548-554.
- Cornsweet, T.N., & Crane, H.D. (1973). Training the visual accommodative system. Vision Research, 13, 713-715.
- Haber, R.N., & Hershenson, M. (1980). The Psychology of Visual Perception (Second Edition). New York: Holt, Rinehart and Winston.
- Lanyon, R.I., & Giddings, J.W. (1974). Psychological approaches to myopia: A review. American Journal of Optometry and Physiological Optics, 51, 271-281.
- Provine, R.R., & Enoch, J.M. (1975). On voluntary ocular accommodation. Perception and Psychophysics, 17, 209-212.
- Renshaw, S. (1945). Tachistoscope in visual diagnosis and training. Optometric Weekly, 36, 1189.
- Ricci, J.A., & Collins, F.L. (1981). The role of psychology in the health care of ophthalmological disorders. Clinical Psychology Review, 1, 431-451.
- Trachtman, J.N. (1978). Biofeedback of accommodation to reduce functional myopia: A case report. American Journal of Optometry and Physiological Optics, 55, 400-406.
- Woods, A.C. (1946). Report from the Wilmer Institute on the result obtained in the treatment of myopia by visual training. American Journal of Ophthalmology, 29, 28-57.
- Young, F.A., Baldwin, W.A., Leary, G.A., West, D.C., Box, R.A., Harris, E., & Johnson, C. (1969). The transmission of refractive errors within Eskimo families. American Journal of Optometry and Archives of American Academy of Optometry, 46, 676-685.



Effects of Image Update Rate on Target Identification Range

Elizabeth L. Martin, PhD
Armstrong Laboratory
Aircrew Training Research Division
Williams AFB, AZ

Julie Mapes Lindholm, PhD
University of Dayton Research Institute
P. O. Box 2020
Higley, AZ

Abstract

In prior research, moving form perception tended to be nonveridical when the update rate of an image generator was less than the refresh rate of the display device. In this experiment, we examined the effect of update rate on the range at which pilots could identify moving aircraft models. The results indicate that target identification range is greater for a 60 Hz update rate than for a 30 Hz update rate. Methodological issues and flight simulation design implications are discussed.

The visual systems of most flight simulators consist of a raster display of computer-generated images. The image generator (IG) and the display device operate concurrently. While one image is being presented, the next image is being calculated.

To create a raster-display image, electron beams trace a fixed pattern of horizontal lines, from left to right and from top to bottom. Most displays have a 60-Hz refresh (vertical scan) rate. If the raster is noninterlaced, all of the horizontal lines are scanned during each refresh period. If the raster is interlaced, the odd and even numbered lines are scanned during successive refresh periods.

Each digital image represents the visual world at a particular point in time. Although most systems change the digital image that is controlling the display only between vertical scans, the image update rate is not necessarily equal to the refresh rate of the display device. An IG coupled with a 60-Hz interlaced display can update once every field (60 Hz) or once every frame (30 Hz). Similarly, an IG with a noninterlaced display can update every frame (60 Hz) or every other frame (30 Hz). Some IGs update at less than 30 Hz or respond to overload problems by reducing their nominal rate. Whenever the update rate is less than the refresh rate, the last image is displayed repeatedly until computation of the next image is completed.

In such systems, relative motion of an object is portrayed by appropriate differences over frames or fields in the size, shape, and location of the object's spatial representation. The

exact sequence of images for a given object trajectory depends upon the update rate of the IG and the refresh pattern and refresh rate of the display device.

Lindholm (1992) examined the effects of image update rate and display interlacing on the perception of a small geometric form moving horizontally at a constant virtual velocity. She found that form perception tended to be nonveridical when the update rate was less than the refresh rate. The apparent form of the target was a composite of the two (or more) target representations which were displayed at each location, with the earlier representation spatially advanced by the distance the target would have traveled during the interdisplay interval. In other words, the perceptual system converted temporal intervals into spatial intervals.

In the present experiment, we examined the effects of image update rate (30 Hz vs. 60 Hz) on the range at which pilots could identify models of air-craft with which they were already familiar: F-15, F-16, MIG-29.

Method

Subjects

Twelve fighter pilots served as observers. Their ages ranged from 28 to 46 years. Prior to participation, each observer was informed of the general purpose of the research.

Apparatus and Visual Displays

The motion sequences were generated by General Electric's Advanced Visual Technology System (AVTS), a Research and Development system similar to the COMPU-SCENE IV. AVTS has an addressability of 985(V) x 1000(H). It supports an interlaced display device. When AVTS is set to operate at a 60-Hz update rate, the digital images for the two fields of a frame are based on different samples of the data base. When it is set to operate at a 30-Hz update rate, the images for the two fields are based on the same sample.

A Barcographics 800 CRT projector was used to display the images on a 6 foot (1.8 m) flat screen. The observer sat 3 feet (.9 m) from the screen with his head position maintained by a chin rest. A two-button response box was used to initiate a series of trials and to indicate aircraft identity. The room was darkened during testing.

Three-dimensional, all-white (6 fL) models of an F-15, F-16, and MIG-29 were created for this experiment. Planform and side views were presented against a blue back-ground (.4 fL). The stimulus set for a given test was one of the three pairs at one viewing angle. Each test trial consisted of a 1-sec horizontal motion sequence, at a constant simulated distance, with the target size and "velocity" (spatial interval between successive locations) determined by the range and air speed of the aircraft. A new trial began 2 seconds after the observer responded.

Design

The observer's task for each set of test trials was to indicate which of the two designated aircraft had been presented. A transformed up-down staircase procedure (Levitt, 1971) provided an estimate of the range at which the observer could perform with 79% accuracy (X_{79}). The target's range for the first trial of each staircase was 2025 ft (1/3 nm). The range on subsequent trials depended upon the response history. Every time the observer made an identification error, the range was decreased. Three correct responses resulted in an increase in range. The magnitude of the step size (change in range) was reduced during the course of a test. The initial step size was .1 log units. The step size was reduced to .05 log units after the first run (sequence of trials in which the changes in range were all in one direction) and to .025 log units after the third run. A staircase was terminated after 5 runs or 60 trials, whichever occurred first.

With the 90 degree field of view, the minimum range at which a full 1-second, 800-knot motion sequence could be displayed was 891 ft. This range was set as a minimum for both speeds. Thus, if the next step in a staircase would bring the aircraft closer than 891 ft, the range was not decreased. Trials were presented at the range just greater than the minimum until the observer was correct three times in succession.

Each observer was tested with all 24 combinations of 2 update rates (30 Hz and 60 Hz), 2 aircraft views (planform and side), 3 aircraft pairs (F-15/ F-16, F-16/MIG-29, F-15/MIG-29), and 2 speeds (400 knots and 800 knots). The testing was distributed across two sessions. Half of the observers were tested with the planform view on the first session and the side view on the second session. Half of the observers were tested in the reverse order. Within a session, trials were blocked according to update rate. The order in which the update rates were presented was counterbalanced across observers and sessions. The order of pair presentation was latinized. Trials for the two speeds for a given update rate x aircraft pair x aircraft view combination were randomly interleaved. The aircraft and direction of motion (left to right or right to left) were chosen randomly on each trial.

Procedure

To ensure that the observer could identify the spatial representations of the aircraft models at a variety of ranges, each set of trials began with a sequence of static presentations. The two aircraft models for that set were presented side-by-side at 500, 1000, 2000, and 4000 feet, respectively. The timing of this cycle, which was repeated twice, was under the control of the observer. Each of the 2 aircraft was then presented for identification 4 times at each of the 3 closer ranges. The order of the six aircraft x range combinations was random within each set of 6 trials. This entire procedure was repeated until the observer made less than 3 identification errors during the 24 identification trials.

The test trials were initiated after the observer had met the identification criterion. Observers were given the opportunity to take a break between sets of trials. A typical testing session was completed in about 45 min.

Results

The log X_{79} estimates for each of the two aircraft views were subjected to a 5-way analysis of variance with two grouping factors (session and update-rate order) and three within-subjects factors (update rate, aircraft pair, and velocity). The alpha level was set equal to .01.

A three-way (update rate x aircraft pair x velocity) repeated measures analysis of variance was conducted on the log X_{79} estimates for each of the two aircraft views ($\alpha = .01$). The results of these two analyses were quite consistent. For both aircraft views, the mean estimate for the 60-Hz update rate was significantly larger than the mean estimate for the 30-Hz update rate: $F(1,8) = 89.09$, $p < .0001$ and $F(1,8) = 190.07$, $p < .0001$, for the planform and side views, respectively (see Table 1). The aircraft-pair effect was also significant in both the planform-view analysis, $F(2,16) = 12.90$, $p < .0005$, and the side-view analysis, $F(2,16) = 11.02$, $p < .001$, although the means were ordered differently in the two cases. The only other effect which reached statistical significance was the update rate x velocity interaction in the side view analysis, $F(1,8) = 16.50$, $p < .005$: Whereas velocity had no effect when the update rate was 60 Hz, the log X_{79} estimates were greater for 400 knots than for 800 knots when the update rate was 30 Hz.

TABLE 1
Geometric Mean Identification Range (ft)

Aircraft View	Update Rate	
	60 Hz	30 Hz
Planform	3100	2181
Side	3896	1885

Discussion

The results of this experiment demonstrate that moving targets will tend to be identifiable at a greater range on a visual system with a 60-Hz update rate than on a visual system with a 30-Hz update rate. The update effect held for two views of three aircraft pairs moving at two different velocities.

This finding has important implications for the design of flight simulators intended to support tactical air-combat training. Traditional engineering approaches have emphasized display resolution and IG address-ability. This study suggests that the IG update rate is also an important design factor. Most high performance visual systems can operate at 60 Hz, usually by reducing scene content. Low cost systems typically operate at 15 Hz or 30 Hz. We would recommend a 60-Hz update rate for any flight simulator intended to support air-to-air combat training. ✓

Although target identification is undoubtedly more difficult with a 30-Hz update rate than with a 60-Hz update rate, the X_{79} values we obtained are probably not good estimates of absolute identification ranges. Identification accuracy always varies with the size of the stimulus set and the similarity of the alternatives. With the procedure used here, only two alternatives were presented during any one set of trials. In addition, the staircase procedure assumes that the expected proportion of a correct response is a monotonic function of the stimulus level (range, in this case). ~~Although we do not have any data to the contrary, a decrease in range was coupled by an increase in image velocity.~~ At the minimum range, an aircraft moving at 800 knots would have traversed the 90-degree field of view in 1 second. It is unlikely that the observers were able to accurately track the target at some of the closer ranges. When a decrease in range reduced the observer's ability to track the target, it may well not have increased the probability of a correct identification response. Finally, the observed reversal values were often highly variable for a given observer. This variability suggests a change in discrimination criteria over time. Informal comments by some of the pilots substantiate this view. ✓

It should be emphasized that the superiority of a 60-Hz update rate presupposes a 60-Hz display device (the current industry standard). According to the analysis of Lindholm (1992), nonveridical form perception will occur when the update rate of the IG is less than the refresh rate of the display device. Thus, if a 120-Hz display were used, a 120-Hz update rate would be necessary. Furthermore, one should consider that research accomplished to date has been limited to the study of the effects of update rate on form perception. Additional research is needed to determine whether update rate has an equally profound impact on the perception of other types of dynamic scenes, such as low-altitude flight imagery. Researchers at AL/HRA are currently investigating the effects of IG update rate on the apparent velocity of textured terrain at low altitudes.

References

- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. Journal of the Acoustical Society of America, 49, 467-477.
- Lindholm, J. M. (January 1992). Temporal and spatial factors affecting the perception of computer-generated imagery (AL-TR-1991-0140). Williams AFB, AZ: Armstrong Laboratory, Aircrew Training Research Division.

AD-P006 950



Individual Differences in a Dual Task Context

Christopher R. Rate
Anthony J. Aretz, Ph.D.
US Air Force Academy

92-17683



Abstract

This paper examines the effects of individual differences on spatial cognition. Twelve subjects, of either high spatial or verbal abilities, were examined while performing a mental rotation task under single and dual task conditions. The results indicate that mental rotation was used by both groups in the single task condition and the verbal group used an alternate strategy in the dual task condition.

Aretz (1988, 1989, 1991) suggests that a pilot maintains navigational awareness through the constant mental alignment of two frames of reference: the ego centered reference frame (ERF) and the world centered reference frame (WRF) (see also Wickens, Aretz, and Harwood, 1989). The ERF is the pilot's view from the cockpit, and the WRF is a typical map with north at the top. Previous studies (Aretz, 1988, 1989, 1991) demonstrate that mental rotation seems to be required to maintain a congruent alignment between the ERF and WRF.

Even though navigation is a primary concern for a pilot, it is not the only task that must be accomplished. Pilots must divide their attention among multiple tasks, some that are spatial in nature (e.g., flight control) and others that are verbal (e.g., communications). One purpose of this experiment was to investigate possible mental rotation performance decrements while performing a concurrent verbal or spatial secondary task. It was hypothesized that more interference would exist between the two spatial tasks than between the spatial and verbal task. In contrast, Williams and Wickens (1991) found equivalent mental rotation performance decrements during concurrent performance with both spatial and verbal secondary tasks.

The difference between the Williams and Wickens (1991) experiment and the present one is that we chose to take an individual differences approach. According to Wickens and Weingartner (1985), high spatial ability subjects may differ from high verbal ability subjects in either the efficiency or the total capacity of their spatial resources. If spatial subjects are more efficient in their use of spatial resources, then smaller dual task performance decrements would be expected for the spatial subjects than verbal subjects when two concurrent spatial tasks are performed. This is because the spatial subjects require less resources for the performance of a spatial task and would suffer less disruption by a concurrent spatial side task. However, if the spatial subjects have more spatial resources than the verbal subjects, then there should be equal performance decrements for both groups, except the spatial subjects' overall performance might be better.

Two previous studies by Aretz (1988, 1989) indicated that there might be a strong influence of individual differences on the performance of mental rotation in a navigation context. The experiments showed the possible existence of systematic differences among the subjects' cognitive abilities in terms of spatial awareness. This is the main emphasis of this paper. The hypothesis of this experiment was that there may be possible differences between high spatial and high verbal ability subjects (from this point, the two groups will be referred to as spatial or verbal) in terms of their cognitive resources or performance strategies. It was believed that the spatial subjects might either have more spatial resources, or be more efficient in using them, than the verbal subjects. If this difference exists, it should result in performance discrepancies between the spatial and verbal subjects while performing spatial and verbal tasks under single and dual task conditions.

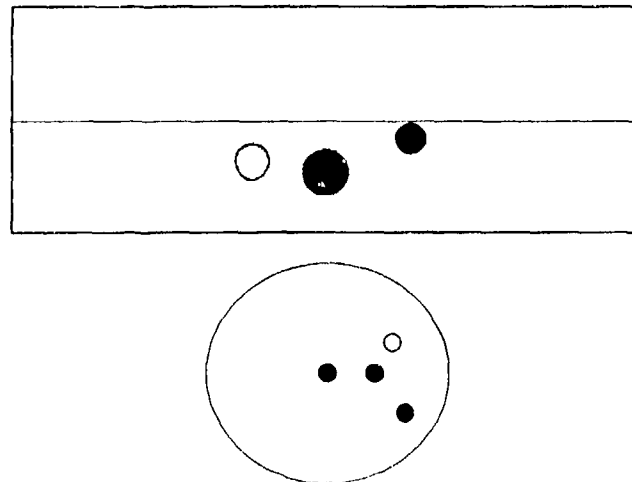
Method

Subjects

Twelve subjects were selected from a pool of 46 freshman volunteers attending the United States Air Force Academy. The subjects were selected based on their standardized scores from a battery of six (three verbal and three spatial) pencil and paper cognitive abilities tests taken from the Educational Testing Service's Cognitive Abilities Test Battery (Erkstrom, French, Harman, & Dermen, 1976). The three verbal tests included the word beginnings, advanced vocabulary, and controlled associations tests. The three spatial tests included the cube comparisons, paper folding, and card rotations tests. Each subject was selected and placed in either the verbal or spatial group (six in each) by scoring in the upper quartile of their group's ability of emphasis and the lower half of the other. All the subjects reported they were pilot qualified or expected to be pilot qualified.

Apparatus

The experiment was conducted in a laboratory room to avoid distractions to the subjects. The navigation task was presented on a standard Zenith 248 computer with a Texas Instruments speech board. The navigation stimuli were presented in the same manner as Williams and Wickens (1991). Circular symbolic map displays with two to three landmarks were presented in the bottom half of the CRT screen (see Figure 1). The top half of the screen contained a three dimensional symbolic representation of the forward view of the world. The color of each circle in the map matched the color of its corresponding sphere in the forward view. All the colored circles were located within a single 90° quadrant of the map. A subject's task was to detect an agreement (match) or displacement (mismatch)



in the location of one of the world's spheres based on its expected location as indicated on the map.

Figure 1. Line drawing of the map comparison stimuli.

Secondary Tasks

This experiment used two different secondary tasks: a verbal digit subtraction task and a spatial grid mapping task. The digit subtraction task was identical to that used by Williams and Wickens (1991). The stimuli for the digit task were presented using predetermined values synthesized by a Texas Instrument speechboard over headphones. A single digit was presented every 4 seconds. The subject's task was to compute the absolute difference between the last two numbers presented, say the value aloud, and remember the last digit presented over the headphones. For example, if the first digit presented was "6", and the next was "4", the subject would say "2" (i.e., $|6-4| = 2$). The subject's

responses were recorded by an experimenter on paper as well as on an audio cassette. The dependent variable for this task was response accuracy.

Since there were possible visual modality interference effects in the Williams and Wickens (1991) data, a different spatial secondary task was selected for this experiment. The task selected was a modified version of the spatial grid task used by Brooks (1968). For this task, subjects were told to imagine a 4x4 grid. They were told to always imagine their starting location as the second square of the second row. Subjects were then given seven directions through the grid at an interval of 4 seconds, again using the speechboard and headphones. The directions were either "right", "left", "up", or "down". After a set of directions were completed, instructions were displayed on the CRT informing the subject to trace the path they had memorized on a pad of paper containing blank 4x4 grids, beginning with their starting position (indicated by a black dot on the paper). The subjects pressed a spacebar to continue with the next set of directions when they were finished. The dependent variable for this task was also response accuracy.

Procedures

The subjects completed one training and one data collection session. Each session lasted approximately 50 minutes, occurred on different days, and both were completed within one week of each other. The training session was identical to the data collection session except that detailed instructions were provided for each task. The order of the tasks was counterbalanced for each subject under the constraint that all the single task conditions preceded the dual task conditions. Following the completion of each condition, the subjects were asked to indicate their perceived difficulty of the condition on a workload rating scale ranging from 0 (very easy) to 100 (very difficult).

For the map comparison trials, subjects indicated their response by using their right hand to press either the "1" or "3" key on the numeric keypad of the computer's keyboard. A "1" indicated a match and a "3" indicated a mismatch. The subjects were instructed to make their response as quickly and accurately as possible. Approximately two seconds separated each trial and subjects received error feedback on the CRT immediately following their response (CORRECT or INCORRECT). Both reaction time and error data were recorded by the computer.

Results

The data from this experiment were analyzed using a mixed model ANOVA. The results are presented in three parts: primary task reaction times, secondary task accuracies, and subjective workload ratings.

As shown in Figure 2, there was a significant main effect for angle, $F(3,30)=15.28$, $p<.001$. As the angle increased from 0° to 180° there was an increasing linear trend in reaction time, as well as a decreasing trend from 180° to 360°. These linear trends indicate that mental rotation was being used by the subjects to perform the map comparison task. Marginal significance was found for the condition main effect, $F(2,20)=3.04$, $p=.07$, where single task map comparisons were faster than the maps plus the grids, which were in turn faster than the maps plus the digits. There was no significant effect for subject type or subject type by condition interaction, however, there was a difference in accuracy. The spatial subjects had an error rate of 17%, while the verbal subjects had an error rate of 24.5%.

Even though a main effect was not found for subject type, Figure 2 also shows a marginally significant three way interaction among subject type, angle, and condition, $F(6,60)=1.94$, $p=.088$. This interaction indicates that the spatial subjects continued to use mental rotation for the map comparisons across all loading conditions, while the verbal subjects seemed only to use mental rotation in the single task condition (i.e., their RTs were relatively constant across rotation angle in the dual task conditions). The error rates did not indicate that this interaction was a result of a speed accuracy trade-off.

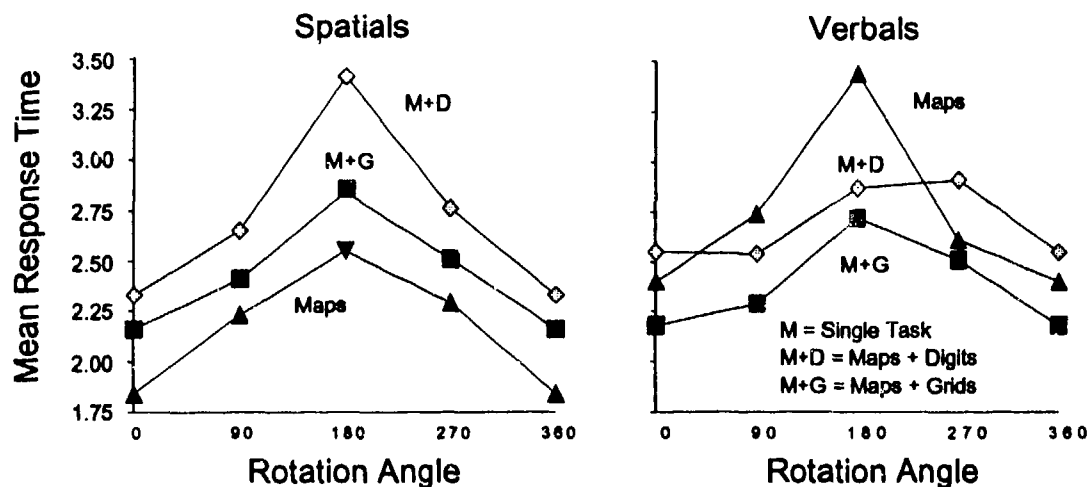


Figure 2. Three way interaction among subject type, condition, and rotation angle.

The secondary task data showed a main effect of condition on both the grid and digit performance, $F(1,10)=4.21$, $p=.067$ and $F(1,10)=14.95$, $p<.005$, respectively. The corresponding error rates were: single task grids (99%), dual task grids (95%), single task digits (87%), and dual task digits (68%). This pattern was similar for both groups of subjects.

Analysis of the workload ratings indicated there were main effects for subject type and condition, $F(1,9)=8.28$, $p<.05$ and $F(4,36)=15.43$, $p<.001$, respectively. More specifically, the verbals reported a higher workload ratings than the spatial, 60.3 and 42.8 respectively. The workload ratings for the five different conditions were: maps (45.4), maps + digits (76.1), maps + grids (69.3), digits (48.1), and grids (24.4). There was not a significant interaction between subject type and workload.

Discussion

The results of this experiment support previous findings by Aretz (1988, 1989) that mental rotation is involved in some of the cognitive operations for aircraft navigation. As figure 2 shows, mental rotation was used by both spatial and verbals in the single task condition, with the spatial being faster. However, under dual task conditions there seems to be a strategy change for the verbals but not the spatial. In the dual task conditions, the spatial continued to use mental rotation (as indicated by the linear trends), while the absence of linear trends for the verbals indicates that the verbals were using an alternative strategy other than mental rotation to perform the task. In fact, the verbals were faster than the spatial in the dual task conditions, but at the cost of a higher error rate.

These data support the hypothesis that the spatial have more spatial resources than the verbals and are not necessarily more efficient (e.g., a higher capacity visual-spatial scratch pad). Because they have more spatial resources, the spatial could delay the onset of the mental rotation task until they were at a convenient stopping point of the secondary task, then proceed uninterrupted with mental rotation. This strategy is evident in the data in that task loading only increased the intercept of the mental rotation function and had little effect on the slope.

The verbals' data indicate an entirely different strategy. Although the verbals used mental rotation in the single task condition, the absence of the linear trends in the dual task conditions indicate an abandonment of mental rotation. In fact, performance improved in the dual task conditions. It seems the verbals abandoned a spatial strategy, presumably because they have fewer spatial resources, in favor of a verbal or analytic strategy that capitalized more on their cognitive strengths. Unfortunately, the

design of this experiment does not allow for the specification of the actual strategy that resulted in the verbals improved performance in the dual task conditions.

Conclusions

The purpose of this experiment was to determine if individual differences influence the cognitive processing strategies used by subjects in a simulated spatial navigation task. The results show that individual differences do seem to play a key role. In the single task condition, both the spatial and verbals use mental rotation to perform the map comparison task, but the spatial were faster. However, in the dual task conditions, the verbals use some other strategy besides mental rotation that results in overall better performance than the spatial. These data suggest the spatial had more spatial resources than the verbals, which allowed for the continued use of a spatial strategy in all task loading conditions. This conclusion is consistent with the findings of Wickens and Weingartner (1985). Further research is required to specify the alternative strategy used by the verbals.

References

- Aretz, A. J. (1988). A model of electronic map interpretation. Proceedings of the 32nd Annual Meeting of the Human Factors Society, (pp. 8-12). Santa Monica, CA: Human Factors Society.
- Aretz, A. J. (1989). Spatial cognition and navigation. Proceedings of the 33rd Annual Meeting of the Human Factors Society, (pp. 89-93). Santa Monica, CA: Human Factors Society.
- Aretz, A. J. (1991). The design of electronic map displays. Human Factors, 33, 85-101.
- Brooks, L. (1968). Spatial and verbal components of the act of recall. Canadian Journal of Psychology, 22, 349-368.
- Wickens, C. D. & Weingartner, A. (1985). Process control monitoring: The effect of spatial and verbal ability and concurrent task demand. In R. E. Eberts & C. G. Eberts (Eds.), Trends in Ergonomics/Human Factors II, (pp. 25-32). New York: Elsevier Science Publishers.
- Wickens, C. D., Aretz, A. J., & Harwood, K. (1989). Frame of reference for electronic maps: The relevance of spatial cognition, mental rotation, and componential analysis. Proceedings of The Fifth International Symposium on Aviation Psychology, (pp. 245-250). Columbus, OH: Ohio State University.
- Williams, H. P. & Wickens, C. D. (1991). Electronic map interpretation in a dual-task context. Proceedings of the 35th Annual Meeting of the Human Factors Society, (pp. 92-96). Santa Monica, CA: Human Factors Society.



Illusions of Unique Invulnerability:
Implications for Selection of Military Personnel for Risky or Hazardous Duties

Captain Kristen Morton Vance
United States Air Force Academy

Abstract

People tend to maintain certain positive illusions about themselves and their futures that may be adaptive in buffering their self-esteem and feelings of efficacy from the effects of negative or threatening feedback. One of these beliefs, the illusion of unique invulnerability, is the expectation that others will be the victims of misfortune and negative events more so than oneself. One possible implication of holding this belief is that if a false sense of security is fostered, actual vulnerability to experiencing negative events that one has control over might be increased if self-protective behaviors are decreased. Alternately, people may feel uniquely invulnerable because they practice precautionary behaviors. The purpose of this study was to examine the nature of the relationship between military cadets' beliefs in unique invulnerability and assumption of risk in behavior while rappelling.

A large component of physical health and well-being is behavioral. Preventive health behavior is frequently urged to protect people from disease, accidents, and environmental hazards. Engaging in risky behaviors also impacts heavily on health. For example, smoking, driving while intoxicated or while not wearing a seat belt, and failing to practice "safe sex" can have drastic health consequences. This can be especially important in military environments where many jobs contain elements of risk (e.g., Air Force fighter pilot, Army airborne ranger, Navy SEAL). Thus, understanding the psychological mechanisms of risky behavior has important implications for health and safety outcomes for both civilian and military populations.

One way of understanding risky behavior is with respect to people's unrealistic optimism about future life events. There is widespread evidence in the literature that normal people believe their futures will be better than the average person's, that good things are more likely to happen to them, and that negative life events are more likely to happen to others (reviewed in Taylor & Brown, 1988). Essentially, people tend to think they are relatively invulnerable, and expect others will be the victims of misfortune more so than themselves. One possible implication is that holding this belief strongly could lead to increased or unnecessary risk-taking. By lulling a person into a false sense of security, objective vulnerability may be increased if self-protective, precautionary, preventive behaviors are decreased. For example, a fighter pilot who underestimates the risk of personal danger or overestimates his or her physical tolerances might purposely engage in an unnecessarily high-g maneuver, lose consciousness, and crash the aircraft. On the other hand, if people feel relatively invulnerable because they take precautions (e.g., don't smoke, wear seat belts, use condoms), then this belief may prove to be related to decreased risk-taking.

The notion that this illusion of unique invulnerability may have behavioral consequences has received limited empirical examination. In a national survey of new-car buyers, Robertson (1977) found that 40% of respondents felt their chances of being injured or killed in a car crash were less than for "people like themselves". However, no statistically significant associations were found between perceived vulnerability and buyer willingness to pay for increased crash protection.

Burger & Burns (1988) reported that female undergraduates estimated their own likelihood of experiencing an unwanted pregnancy at 9%, while estimating the likelihood for average American females at 43%. Further, the stronger this illusion of unique invulnerability, the less likely subjects were to use effective contraception when having intercourse, based on self-reports.

Finally, Weinstein, Sandman, and Roberts (1991) tested the hypothesis that perceptions of personal susceptibility are important in decisions to test one's home for radioactive radon gas. Perceptions of personal susceptibility predicted both subsequent radon test kit orders and testing intentions, but not the actual behavior of conducting the radon testing.

Purpose of Current Research

Evidence from social cognition research clearly suggests that people appear to have the capacity to distort reality in a direction that promotes an optimistic view of the future. Despite wide empirical support for the existence of this cognitive phenomenon, the link between perceived invulnerability and assumption of risk in subsequent behavior is not well established. The current data that address the possible impact of beliefs of invulnerability on behavior are inconsistent, and rely largely on self-report measures.

The purpose of the current research was to pursue the link between unique invulnerability and behavior, and to address the methodological limitation mentioned above. The primary goal was to establish empirically, via direct observation of behavior, whether holding beliefs of relative invulnerability relates to behavior and affects assumption of risk while engaging in an inherently risky activity, rappelling. Originally associated with mountaineering, rappelling serves as an end in itself for a few enthusiasts who devote leisure time to the descent of practice cliffs, buildings, and other tall structures. However, experienced rock climbers and mountaineers often view rappelling as a risky endeavor. One rock climbing handbook calls this "the most dangerous maneuver in mountaineering", to be used only as a last resort (Gregory, 1989). Based on the relationship described by Burger and Burns (1988), it was hypothesized that subjects who perceived themselves as relatively more invulnerable would behave more riskily (less safely) while rappelling than subjects who perceived themselves as relatively less invulnerable. This hypothesis was admittedly speculative because, as described earlier, the possibility that relative invulnerability may be related to decreased risk-taking seemed equally feasible.

Method

Subjects

Subjects were 157 Reserve Officer Training Corps (ROTC) cadets, male and female, ages 17-20, entering their freshman year and undergoing summer military training at Virginia Polytechnic Institute and State University. Rappelling training was mandatory for all incoming ROTC cadets. However, participation in this study was on a strictly voluntary basis. No cadets declined to participate in the study.

Materials

Unique invulnerability scale. The invulnerability scale measured perceived invulnerability to negative life events in general (e.g., "Having a heart attack"), and to specific negative events associated with rappelling (e.g., "Spraining an ankle or breaking a bone"). The scale consisted of two parts: self ratings, in which subjects indicated the likelihood on a 7-point Likert scale (1="Not at all likely" to 7="Extremely likely") that they would experience certain negative events, and other ratings, in which subjects indicated the likelihood that the "average ROTC cadet of your sex at Virginia Tech" would experience the same ten events. The two parts were counterbalanced during administration. The scale was analyzed via difference scores, calculated by subtracting self ratings for each event from the corresponding other ratings. A positive score reflected perceived relative invulnerability towards that event. An invulnerability index was calculated by taking the mean self-other difference score across events. The invulnerability scale was created specifically for use in this study, and therefore had not been administered previously. However, the general negative life events were adapted from Weinstein (1984) and Perloff and Fetzer (1986) whose research showed that they reliably evoke perceptions of invulnerability in college students.

Behavioral observation. The following resources were used to observe and measure subjects' performance during rappelling training: two videocameras and cameramen; eight research assistants; and data collection sheets. The research assistants were trained on the data recording procedures using a videotape which was developed in cooperation with the ROTC staff.

The tape showed a variety of rappels, similar to what was encountered during data collection. Anchors for all the behavioral measures were represented.

A post hoc reliability check of the rappelling data (number of bounds) was performed to assess whether observations made in real time matched observations made of the videotaped jumps. A research assistant and the experimenter independently observed 40 videotaped rappels (15% of the total sample). Results indicated that the experimenter agreed with 90% of the real-time observations, the assistant with 87% of the real-time observations, and the experimenter with 89% of the research assistant's observations. Thus, interrater reliability in the real-time recording of the behavioral data appeared to be at an acceptable level.

Procedure

Administration of the pre-measure. Five days before the rappelling training, in a single group session, subjects were asked to participate in a research project concerned with the impact of people's beliefs about themselves on military training. The unique invulnerability scale was administered to assess subjects' beliefs regarding invulnerability prior to rappelling.

Observation during rappelling training. Subjects were observed as they accomplished a total of one to three rappels each off a 20 foot and/or a 40 foot tower. The rappelling session was conducted by ROTC upperclass cadets and staff at the rappelling tower on campus. Several behavioral measures were recorded on data sheets, and the entire rappelling session was videotaped to allow post-hoc reliability checks of the behavioral data. During the training, research assistants recorded behavioral data from a variety of positions: one assistant was located on top of each tower (the 20 and 40 foot), and two at the base of each tower.

Behavioral measures. The behavior of interest, rappelling, involves descending a tower by lowering oneself down a rope attached to the top of the tower. The descent is properly accomplished by pushing or "bounding" out from the tower, falling several feet, braking one's progress by tightening on the rope, landing back on the wall, and continuing this series of steps until reaching the ground. The key to rappelling safely lies in making every move slowly and carefully. Rappellers should go down the rope smoothly, avoiding the "commando-jump" style of the movies. Such antics strain the anchor and the rope, and may cause them to fail. One also risks spraining or breaking an ankle by swinging back into the tower wall too quickly or forcefully. Long leaps, fast drops, and sudden stops should be avoided (Lyman, 1975).

Three behavioral measures were used to tap assumption of risk. The first was number of bounds to descend the tower. The number of bounds variable was characterized a priori in the following way. For rappels off the 20 foot tower, less than two bounds during descent was defined as "risky", two to three bounds as "average", and more than three bounds as "conservative". For rappels off the 40 foot tower, less than three bounds was considered "risky", three to four bounds "average", and five or more bounds "conservative". These definitions were arrived at through discussions with the rappelling instructor cadre regarding what constituted reasonable (average) performance for a naive subject group, and what characterized unsafe (risky) and ultra-safe (conservative) behavior.

The second behavioral measure was decision to rappel. Though it was mandatory to attend the training, cadets were not required to rappel. Rappelling was defined as involving more risk than not rappelling.

The third behavioral measure, instructor rating, was related to risk via overall performance. It was difficult to disentangle unsafe behavior from overall jump performance; generally a "poor" jump was also considered an unsafe jump. Thus, instructor rating was a subjective rating of the quality and safety of each rappel. Instructors evaluated each rappel as either a 1 ("excellent"), a 2 ("average"), or a 3 ("poor"), based upon a standardized set of criteria.

Results

Due to time constraints beyond their control, subjects varied widely in the number of jumps taken, which ranged from one to three jumps total. Therefore, behavioral measures were analyzed separately for first jumps, second jumps, and third jumps, and for jumps off the 20 foot and 40 foot towers. Also, three groups of subjects were considered: subjects who could not rappel due to time constraints ($n=28$), subjects who chose not to rappel ($n=13$), and subjects who did rappel at least once ($n=116$).

Instructor Rating and Number of Bounds

Data for first rappels off the 20 foot tower revealed that relative invulnerability was negatively correlated with instructor rating, $r=-.22$, $p<.05$. That is, the more invulnerable the subject felt, the better rating of jump quality he or she received. Invulnerability was not reliably correlated with number of bounds.

For first rappels off the 40 foot tower, invulnerability was not significantly correlated with ratings of jump quality. Relative invulnerability was negatively correlated with number of bounds, $r=-.65$, $p<.01$. Thus, as beliefs in relative invulnerability increased, number of bounds decreased.

For second and third jumps off both the 20 and 40 foot towers, invulnerability was not reliably correlated with instructor rating or number of bounds.

Decision to Rappel

Planned comparisons were conducted for the two between-group contrasts that seemed most relevant to the decision to rappel. First, subjects who could not rappel were compared with subjects who did rappel. Because the former group was functionally a control group, no differences in invulnerability scores were expected between these groups. As predicted, scores on relative invulnerability were not significantly different between these groups, $t(142)=1.22$.

Next, subjects who chose not to rappel were compared with subjects who did rappel. In this comparison, differences between groups in terms of invulnerability scores seemed to be more likely. The specific prediction was that subjects who decided to rappel would have higher invulnerability scores than subjects who decided not to rappel. Relative invulnerability scores did differ significantly between these groups. Subjects who chose not to rappel were reliably different from subjects who rappelled ($M=.56$ versus $M=.96$), $t(127)=-2.05$, $p<.04$, indicating that subjects who elected to rappel held greater beliefs in unique invulnerability than subjects who did not.

Discussion

It was hypothesized that subjects who felt relatively more invulnerable to experiencing negative events would behave more riskily (less safely) than subjects who felt less invulnerable. Upon initial inspection, the data appear to partially support this hypothesis. The decision to rappel, which entailed greater assumption of risk than not rappelling, was related to higher beliefs in relative invulnerability. Also, for first rappels off the 40 foot tower, higher levels of relative invulnerability were related to taking fewer bounds down the tower, a behavior that was operationally defined as involving greater risk than taking more bounds.

However, viewing the data from a different perspective hints at an alternative interpretation. Examination of the mean number of bounds for first rappels off the 40 foot tower ($M=4.41$, $SD=1.46$) indicates that no subject took less than three bounds, which was the operational definition of "risky". Instead, the majority of subjects were in the "average" to "conservative" range with respect to number of bounds. This suggests that although high invulnerability subjects took fewer bounds than low invulnerability subjects, the former were not behaving more riskily but, rather, more competently. Similarly, better instructor ratings of jump quality and safety were associated with greater beliefs in invulnerability, and not with lower invulnerability, as predicted a priori. A possible explanation is that higher levels of invulnerability may have buffered subjects

against the initial fear they felt before the first rappel, allowing them to perform more competent rappels. That high invulnerability was related to both fewer bounds and better instructor ratings seems to indicate that these subjects may have been behaving more competently than low invulnerability subjects, who were behaving more tentatively and conservatively.

Suggesting that number of bounds and instructor rating may be related to competent behavior as opposed to risk-taking is not inconsistent with Taylor and Brown's (1988) theory concerning the functional value of illusions. By buffering people's anxiety in threatening situations, the illusion of unique invulnerability may facilitate more competent performance. In an analysis conducted to explore the utility of such an interpretation, subjects were dichotomized into two groups, high- and low-invulnerability, based on a median split between relative invulnerability scores. For first jumps off the 40 foot tower, subjects low in unique invulnerability bounded more ($M=5.17$, $SD=.75$) than subjects high in invulnerability ($M=4.00$, $SD=1.61$), $t(15)=1.66$, $p<.12$. Considering that three to four bounds on the 40 foot tower was operationally defined as "appropriate", this finding suggests that subjects low in relative invulnerability behaved ultra-conservatively, while subjects high in invulnerability behaved competently.

Application to Personnel Selection for Hazardous Duties

The correlational data obtained in this study provide preliminary support for a link between beliefs in unique invulnerability and the decision to engage in a risky or hazardous activity. It was hypothesized that holding these beliefs might increase one's objective vulnerability and risk by decreasing the perceived need to engage in self-protective behavior. This position was partially supported by data that showed that the decision to engage in a risky activity, rappelling, was related to beliefs in invulnerability.

A possible alternative interpretation of the data emerged as well. It could be, as Taylor and Brown (1988) contend, that holding beliefs in unique invulnerability is adaptive, allowing people to live more happy, caring, productive lives by allowing them an optimistic outlook on the future. The utility of this perspective was tentatively confirmed by evidence that suggested that beliefs in invulnerability were related to more competent performance. The scope and reliability of this possible relationship between unique invulnerability and competence of behavior in risky activities needs to be examined and tested in future research.

In either case, for purposes of selecting military personnel who will be motivated for assignments involving hazardous duties and who will perform competently once assigned, these data tentatively suggest that people with stronger beliefs in unique invulnerability may fit the bill better than people who feel relatively more vulnerable to mishap.

References

- Burger, J. M., & Burns, L. (1988). The illusion of unique invulnerability and the use of effective contraception. *Personality and Social Psychology Bulletin*, 14, 264-270.
- Gregory, J. F. (1989). *Rock sport*. Harrisburg, PA: Stackpole Books.
- Lyman, T. (1975). *The field book of mountaineering and rock climbing*. New York, New York: Winchester Press.
- Perloff, L. S., & Fetzer, B. K. (1986). Self-other judgments and perceived vulnerability to victimization. *Journal of Personality and Social Psychology*, 50, 502-510.
- Robertson, L. S. (1977). Car crashes: Perceived vulnerability and willingness to pay for crash protection. *Journal of Community Health*, 3, 136-141.
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, 103, 193-210.
- Weinstein, N. D. (1984). Why it won't happen to me: Perceptions of risk factors and susceptibility. *Health Psychology*, 3, 431-457.
- Weinstein, N. D., Sandman, P. M., & Roberts, N. E. (1991). Perceived susceptibility and self-protective behavior: A field experiment to encourage home radon testing. *Health Psychology*, 10, 25-33.

Inertia and Adaptation in the Sequential Processing of Information¹Elliot E. Entin
ALPHATECH, Inc.

92-17685



Abstract

The contrast-inertia model of Einhorn and Hogarth (1987) suggests that decisionmakers revise their beliefs differently depending on the order in which they view a sequence of confirming and disconfirming evidence. Results reported in Entin et al. (1989) support this hypothesized order effect. It has been suggested that the order effect may be sensitive to the elapsed time between information presentations. The current experiment replicated the findings of Entin et al. using a different procedure and longer time intervals between information presentations. Results again demonstrate the order effect predicted by the Einhorn and Hogarth theory, and also show that this effect is robust for longer time intervals. We conclude that the order in which pieces of evidence are presented to decisionmakers has a critical effect on their eventual beliefs, sometimes resulting in contradictory opinions as to the likelihood of a threatening event.

Decisionmakers (DMs) do not appear to learn nearly as much from acquired information as the information warrants (Edwards and Phillips, 1964; Peterson and Beach, 1967; Donnel and DuCharme, 1975). There seems to be some inertia associated with a DM's current belief that resists change. Thus, most DMs do not adjust their current position sufficiently given the implications of new information (Tversky and Kahneman, 1974). In recent experiments it appears that the sequential ordering of information and particularly the sequential ordering of confirming and disconfirming information have a differential effect on a DM's current position (Einhorn and Hogarth, 1987; Serfaty et al., 1988; Tolcott et al., 1987).

Einhorn and Hogarth (1987) have described a contrast-inertia theory that predicts that disconfirming evidence will produce a larger adjustment to a DM's position than confirming evidence. This theory suggests that DMs evaluate incoming information as either supporting or refuting their currently held position and thus evaluate evidence as being relevant to only a single hypothesis. A contrast (disconfirming evidence) tends to overcome the DM's inertia to change, making DMs very sensitive to contrasts or discrepancies. The conflict between inertial forces (resistance to change) and adaptive forces (adjustment to new information) coupled with the contrast effect implies that the sequential order in which confirming and disconfirming evidence is received, even if the two pieces of evidence are derived at the same time, will lead to different final beliefs and, perhaps, to different decisions.

To test the hypotheses derived from the Einhorn and Hogarth (1987) theory, Entin et al. (1989) employed an abstraction of a missile warning officer's task using a computer-based simulation. Findings indicated that the order in which pieces of evidence were submitted to subjects had a significant effect on their beliefs, often resulting in contradictory opinions as to the presence or absence of an enemy attack. Subjects were presented with an initial intelligence message and two maps. An attack-likelihood estimate based on collected information was contained in the intelligence message (two levels were used—25% or 75%). The maps then presented patterns of information that either supported or refuted the initial intelligence message. Subjects were told that the two maps were derived from two equally reliable sources at about the same time.

¹ The author would like to express his gratitude to Jean MacMillan and James C. Deckert for their generous editorial assistance. Research was supported by contract DCA100-87-C-0083.

The order in which the information was presented on the two maps affected subjects' ultimate beliefs as predicted by Einhorn and Hogarth's contrast-inertia model. Confirmation followed by disconfirmation (C-D) and disconfirmation followed by confirmation (D-C) elicited vastly different positions from the DMs. When an initial attack likelihood of 25% was provided, subjects experiencing the C-D order switched their beliefs from non-attack to attack after the second map. Subjects experiencing the D-C order first changed their belief to attack, and then back to non-attack. The same patterns were obtained when the initial likelihood of an attack was given as 75%. Recall that an optimal DM applying Bayesian formulations would arrive at identical outcomes for the C-D or D-C orders.

These results have disturbing implications for DMs who are receiving information and trying to formulate an opinion. It appears that the order in which pieces of evidence are considered may outweigh the effects of the information itself. There is some doubt, however, about whether this contrast effect will endure over time. Interestingly, all of the research reported in support of the Einhorn and Hogarth (1987) hypotheses presented pieces of evidence in a very short time period, usually from one or two minutes. Military commanders more typically receive their information over a long period of time. The passage of time between pieces of information may reduce both the inertial effect for the current position and the contrast effect produced by disconfirming information. Thus, a longer time period between the receipt of pieces of information may ameliorate the order effect observed in earlier experiments.

The experiment reported here was designed to test the robustness of the order effect to the passage of time using DMs with military training and a different set of procedures to induce the order effect. In accord with the inertia and adaptation model and previous findings, we hypothesized that subjects would exhibit an order effect even when information was presented over relatively longer periods of time. Specifically, we predicted that subjects would make greater downward revisions in their beliefs after disconfirming evidence than upward revisions in their beliefs after confirming evidence. We expected that this pattern would hold regardless of the initial starting belief.

Method

Overview

Subjects were presented with a written initial intelligence brief followed by two updated messages. The initial intelligence brief described one of four threat events that could affect the subjects during a simulated scenario. An example of one such threat event is the possibility that a terrorist group operating in the Persian Gulf might use high speed patrol crafts in suicide attacks against U.S. ships. Each brief contained the intelligence analyst's likelihood estimate that the threat event would occur. Each brief also requested that the subjects give their likelihood estimate that the described event would occur.

At approximately 30 minutes and 60 minutes after the initial intelligence briefing, subjects received update messages. Each message presented information that either supported or refuted the initial intelligence brief. The two update messages contained information indicating that the evidence in each message had been gathered at about the same time, even though the messages were presented at different times. Subjects estimated the likelihood of the threat event after the first update and again after the second update.

Subjects

Thirty officers enrolled in two operations research courses at the Naval Postgraduate School (NPS) served as subjects. The officers were about equally divided among the Navy, Air Force, and Army. Almost all held a rank between O-1 and O-3.

Independent Variables

Anchor Probability—in accordance with the procedures described in Entin et al. (1989) two anchors or initial probabilities indicating the likelihood that a threat event would occur were employed. The written intelligence brief stated that the intelligence analyst estimated the likelihood that the threat event would occur as either 25% or 75%. Two of the four threat events indicated a 25% likelihood while the other two specified a 75% likelihood.

Confirming and Disconfirming Evidence—the first and second update messages were used to manipulate the sequence of confirming and disconfirming evidence. One initial intelligence brief with a 25% threat likelihood and one with a 75% likelihood were coupled with update messages that first confirmed and then disconfirmed the initial anchor's position. The two remaining intelligence briefings were coupled with update messages that first disconfirmed and then confirmed the initial anchor's position.

Dependent Variable

The primary dependent measures were the three sequential threat likelihood estimates given by the subjects. The likelihood that a threat event would occur was indicated on an eleven-point scale. The scale was anchored at one end by zero and the phrase "no chance of occurrence" and at the other end by 100 and the phrase "almost certain chance of occurrence."

Procedure

Subjects participating in this experiment were part of a larger experiment involving four 90-minute crisis-management scenarios. The four scenarios were simulated using a high-fidelity wargaming facility at the NPS. Each scenario provided an ongoing mission and an attack on U.S. interests that triggered the crisis. The four threat events used in this experiment were conceived and written to be plausible occurrences in any of the four crisis scenarios.

The initial intelligence brief was presented just prior to the onset of the crisis scenario. Subjects read the message and gave their likelihood estimate on the scale provided. The likelihood estimates were collected under the guise that higher command wished to know the likelihood that officers in the field assigned to the threat events. Approximately 30 minutes after the initial intelligence brief, the simulation scenario was paused for data collection and the subjects received the first of the two update messages. The second update message was given to the subjects 30 minutes after the first when the simulation was once again paused for data collection.

Design

The three times at which information was received and probability estimates were made were nested within the two anchor-probability conditions. The two information orders (C-D and D-C) were crossed with the times, and thus two 3×2 repeated-measures within-subjects analysis of variance designs were formed. The four different threat events were counter-balanced across the four different crisis scenarios, and each subject experienced all of the threat events in a different randomized order.

Results

Results for the 75% anchor condition are shown in Fig. 1. As hypothesized, a substantial recency effect was found, substantiated by the significant time step $\times$ information order interaction ($p < .001$). The interaction further indicates that the orders of the confirming and disconfirming information had a differential effect on the way in which subjects revised their initial beliefs.

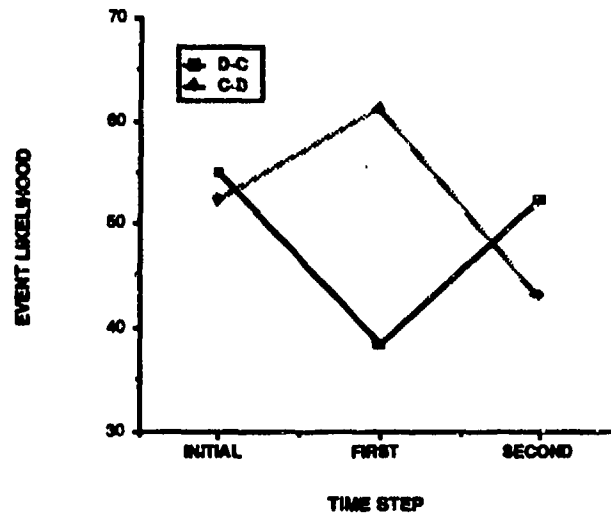


Figure 1. Order Effect for 75% Anchor

Initially, subjects were less certain than the intelligence analyst that threat events would occur. Subjects estimated the likelihood of the threat event at around 55%, substantially less than the 75% estimate provided by the intelligence analyst. Subjects confronted with confirming evidence increased their likelihood estimates by 17%, while subjects presented with disconfirming evidence lowered their likelihood estimates by 30%. As hypothesized by Einhorn and Hogarth (1987) and found by Entin et al. (1989), we see that the negative effect of disconfirming evidence is greater than the positive effect of confirming evidence. DMs are apparently more sensitive to discrepancies than to confirmations.

The second update message presented disconfirming evidence to those who had previously received confirming information and confirming evidence to those who had previously received disconfirming information. In the first case, disconfirming evidence caused subjects to make a substantial downward revision of 30% in their likelihood estimates to a level that is 18% below their initial estimates. In the second case, confirming evidence influenced subjects to revise their current likelihood belief upward by about 37% to a level about even with their original estimates.

Figure 2 shows that the results of the 25% anchor condition mirror those of the 75% anchor condition. A significant time step $\times$ information order interaction ($p < .001$) indicates a strong recency effect and differential effects of confirming and disconfirming evidence on current belief. Subjects' initial estimates were higher than the 25% likelihood that was provided, and, for unknown reasons, the two groups gave significantly different initial likelihoods ($p < .05$). However, subjects responded to the ordering of information (C-D and D-C) in precisely the same manner as in the 75% anchor condition. That is, larger adjustments were made to current beliefs following disconfirming than following confirming evidence.

Discussion

We set out to evaluate the robustness of the inertia-adaptation and contrast effects to the passage of time between the receipt of pieces of evidence. The findings clearly indicate that inertia-adaptation and contrast effects operate over relatively long time periods between the presentation of evidence. The results replicate the earlier findings of Entin et al. (1989) and support Einhorn and Hogarth's (1987) contrast-inertia models.

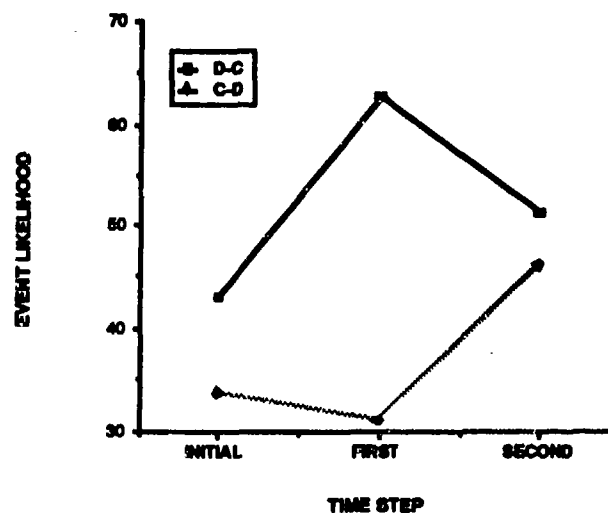


Figure 2. Order Effect for 25% Anchor

The order in which confirming and disconfirming evidence was presented to DMs clearly affected their final beliefs. In some cases different orderings of the same information produced contradictory likelihoods for the same threat event. These results were obtained from trained military DMs, underscoring the importance of making commanders aware of this effect. It is imperative that commanders reach decisions based on the implications of the evidence received, not on the order in which it is received. At minimum, commanders should be made aware of order effects during training and provided with procedures to counter these effects. In addition, some form of decision aid may help commanders and other DMs fuse information in the future so that they do not fall prey to order effects or other deleterious biases.

References

- Donnel, M.L. & DuCharme, W.M. (1975). The effect of Bayesian feedback on learning in an odds estimation task. Organizational Behavior and Human Performance, 14, 305-313.
- Edwards, W. & Phillips, L.D. (1964). Man as transducer for probabilities in Bayesian command and control systems. In M.W. Shelly II & G.L. Bryan (Eds.), Human Judgments and Optimality (pp. 360-401). New York: Wiley, 1964.
- Einhorn, H.J. & Hogarth, R.M. (1987). Adaptation and inertia in belief updating: The contrast-inertia model, Chicago, IL: Technical Report Center for Decision Research, University of Chicago.
- Entin, E.E., Serfaty, D., & Forester, J. (1989). Sequential processing of information from multiple sources. In Proceedings: System, Man, and Cybernetics Society. Cambridge, MA.
- Peterson, C.R. & Beach, L.R. (1967). Man as an intuitive statistician. Psychological Bulletin, 68, 29-46.
- Serfaty, D., Entin, E.E., & Tenney, R.R. (1988). Planning with uncertain and conflicting information. In S.E. Johnson & A.H. Levis (Eds.), Science of Command and Control: Part II. Coping with Complexity (pp. 91-100). Fairfax, VA: AFCEA International Press.
- Tolcott, M.A., Marvin, F.F. & Lehner, P.E. (1987). Effects of early decisions on later judgments in an evolving situation (Technical Report 87-10). Falls Church, VA: Decision Science Consortium, Inc.
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124-1131.



Future Research Applications for a
Job Aiding/Training Conceptual Methodology

Kara B. Worthington
Theodore A. Lamb, Ph.D.
Armstrong Laboratory, Human Resources Directorate

Abstract

This paper addresses research enhancements to the Job-Aiding/Training Allocation Technologies (JATAT) methodology, a conceptual decision support system to assess whether to job aid or to train tasks within a job specialty. Enhancements to the model, in the form of linkage to the Training Decisions System (TDS) and development of a neural network are discussed.

In the face of ever-present budget cuts, a dwindling pool of personnel, and diminishing training resources, DoD needs to develop "smart" means of making the crucial manpower, personnel and training (MPT) decisions which impact the current and future makeup of the services. The Armstrong Laboratory Human Resources Directorate (AL/HR) has developed a means to assist the functional and training communities with these MPT decisions: the Job-Aiding/Training Allocation Technologies (JATAT).

Job Aiding Vs Training

JATAT is a computer-based decision support system to assess which tasks within a job should be job aided, which tasks should be trained, and which tasks require a combination of training and aiding (Zenyuh, Duncan, Rouse, & Lamb, 1991).

The criteria involved with determining the proper mix of training and aiding are quite complex and are illustrated in Figure 1. The knowledge, skills, and abilities (KSAs) needed to perform the required tasks, the job aiding and training requirements associated with the tasks, potential hazards to the human or system, task criticality, resource capacities, and instructional barriers must be assessed to provide a realistic framework to determine the tradeoffs, or alternative training scenarios and their associated costs and benefits (Worthington & Lamb, 1992). The outcome of the tradeoff analyses is the instructional and implementation recommendations in the form of training alone, job aiding alone, or some combination of training and aiding.

Future Research Thrusts

Future research may build upon exploratory developmental work to link JATAT with MPT technologies such as the Training Decisions System (TDS), explore weapon system applications, and expand the prototype DSS.

JATAT is conceptually linked to TDS, which is a computer-based system to help analysts determine what training is required in a specialty, at what point in an airman's career the training should be provided, and what training setting is appropriate, i.e., residence course, field training detachment (FTD), on-the-job training (OJT), or self-paced study. This "what, when, and where" of training allocation will aid in pipeline management, provide

valuable information during utilization and training workshops (U&TWs), and assist functional managers with numerous other MPT decisions. Information such as cost analyses, training capacities and constraints, course lengths, OJT hours, MAJCOM crossflow, changes in the experience/grade level mix, and the effect of bypassing formal technical school can be evaluated for alternative training scenarios prior to implementation.

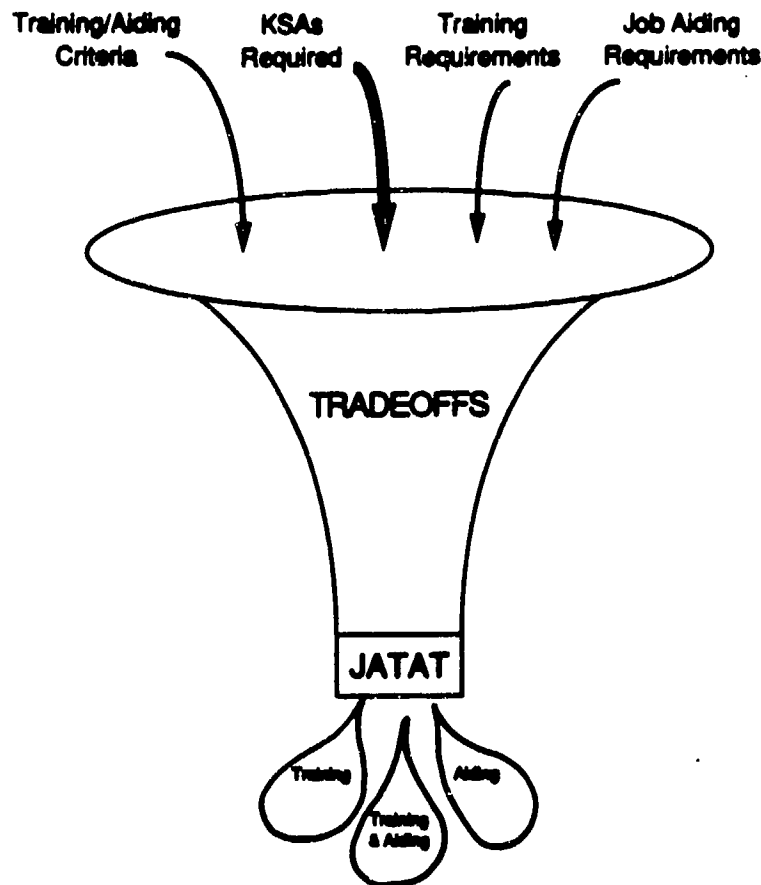


Figure 1. JATAT Funnel

JATAT is termed the front-end analysis of TDS; JATAT's role in determining the proper mix of aiding and training has direct implications for the follow-on decisions of how best to provide the training; i.e., when and where to train the tasks. JATAT and TDS complement each other in using MPT data to highlight the advantages and disadvantages of training outcomes before the options are implemented. Job-aiding where appropriate and providing the proper "what, when, and where" of training maximizes training resources and operational readiness while minimizing expenditures.

While the JATAT and TDS methodologies are proving to be valid predictors within the MPT community (Zenyuh et al., 1991; Mitchell et al., 1990), they also hold much promise for weapon system acquisition. Future research may explore the feasibility of direct weapons systems applications.

Further research may also focus on tasks designed to validate and refine the JATAT model. The model must be subjected to rigorous validation

procedures to include field testing. The prototype DSS must be expanded to incorporate all steps of the JATAT methodology to include a detailed file log of interim and final inputs and decisions.

Artificial Intelligence Enhancements

Applications from the domain of artificial intelligence (AI) can perhaps contribute most to the JATAT model. The Armstrong Laboratory has been designated a center of excellence for artificial intelligence; the role of neural networks is an expanding focus within AI research which has profound implications for developing "smart" computer programs such as JATAT. Configuring the JATAT system as a neural network capable of learning to decipher the vast array of inputs and responding with appropriate training/aiding solutions can catapult the decision-making process into a level of heightened accuracy and efficiency previously unknown.

The architecture of a neural network is composed of artificial neurons interconnected in layers; the numerous interconnections create almost limitless connectivity, and the output of the system can be changed for a given set of inputs so that the system can adapt to learn or be trained in a process of repetitive data sampling to generate its own rules (Robinson, 1990a, 1990b). This process provides the mechanism for making artificial neurons variable, ranging from minus one to plus one, and from positive to negative (Robinson, 1990b).

Assignment of appropriate weights to the information carried by the artificial neurons in the JATAT methodology would be based on the complex integration of all inputs. Weightings could be altered by the neural network to conform to the changing parameters continuously faced by military decision-makers. Changes in the DoD budget; the number of students entering the varied levels of formal training and OJT; the force levels of individual specialties, as well as the overall force level; the equipment and resource capacities and constraints for instruction, general maintenance, and operational missions; and the changing training philosophies resulting from the conversion to composite wings are just a sample of the interacting variables which may be altered at any given point in time. A neural network could determine the complex intricacies of how changes in one input variable could affect other variables and could foresee the long range implications of such interactions. The current JATAT methodology does not incorporate AI principles, but future research thrusts within the Armstrong Laboratory may outfit JATAT with an artificial neural network.

In the current version of the JATAT Decision Support System (DSS), a microcomputer-based demonstrator of the JATAT methodology, the demonstrator utilizes seven models; however, there are a wide variety of methods and models discussed in the literature which could be executed on the DSS. Rouse and Johnson (1989) assessed 29 different procedures. Tools addresses in their technical paper were selected by mapping three general approaches to resolving training/aiding tradeoffs (behavior simulations, performance predictions, and guidelines) against three general sources of information for applying those approaches (judgement, archives, and models).

Evaluating the plethora of available procedures and selecting the appropriate model for use requires a complex process which would be facilitated by artificial intelligence. Enhancing the JATAT methodology to learn the various models, apply the models to the numerous informational inputs associated with evaluating aiding/training tradeoffs, and select the model appropriate for analysis would result in expedient processing of the

complex interactions and an improved analysis capability to apply combinations of models or develop new models. Another adaptation of artificial intelligence to the JATAT methodology is the expansion of the model to use neural networks to develop job aids and training curricula.

Conclusion

The JATAT methodology, whether artificial applications are incorporated or not, provides the functional and training communities with a sophisticated tool with which to more accurately assess and predict the tradeoffs associated with various training-related decisions. In our complex military environment of "doing more and more with less and less," we need to empower our decision-makers with access to the most relevant and up-to-date information at their fingertips; we need to push JATAT into the operational sector to enable functional managers to assess alternative training scenarios prior to implementation in order to maximize training resources while minimizing expenditures.

Future research enhancements in the form of linkage to the Training Decisions System, application to weapon system acquisition, and model updates will facilitate technology transfer from the laboratory to the operational setting. Artificial intelligence applications can enable models to be evaluated singly and in conjunction with each other to adapt existing models and develop new models which more accurately integrate the MPT information for a given specialty. Development of job aids and training curricula can also be facilitated by "smart" programming.

References

- Mitchell, J. L., Vaughan, D. S., Knight, J. R., Rueter, F. H., Fast, J., & Haynes, W. R. (1990, December). Training decisions technology analysis (Report No. F41689-88-D-0251).
- Robinson, C. A., Jr. 1990. Real time expert systems ease production problems. Signal, 44(8): 39-40.
- Robinson, C. A., Jr. 1990. U.S. artificial intelligence is pivotal to scientific aims. Signal, 44(8): 31-35.
- Rouse, W. B., & Johnson, W. B. 1989. Computational approaches for analyzing tradeoffs between training and aiding (Report No. 89-78).
- Worthington, K. B. & Lamb, T. A. (1992). Training tradeoff analyses: A conceptual methodology. Manuscript submitted for publication.
- Zenyuh, J. P., Duncan, P. C., Rouse, W. B., & Lamb, T. A. 1991. A job aiding/training decision process model. Unpublished contractor's report.



Gender Differences: Stress Effects on General Well-Being and Absenteeism

William H. Hendrix
Clemson University
Gail S. Gibson
Alabama A&M University

Abstract

This research investigated gender differences associated with the relationship between antecedents and outcomes (emotional and physical well-being, and absenteeism) of job and life stress. The pattern of relationships between the stress related predictors and criterion variables were similar for males and females. The relationships, predicting absenteeism were, however, more complex for females than those for males.

According to a recent literature review on work stress (Ganster & Schaubroeck, 1991), more than 300 published academic journal articles have dealt with work stress over the past 10 years. Hundreds of additional articles on work stress were also published during this time frame in the popular press as well as a series of books. This level of research and publishing suggests that work stress has been a major area of interest to those performing research as well as to the public in general. Another area receiving attention in organizational research that has been hypothesized as an organizational outcome affected by stress is absenteeism (Hendrix, Steel, Leap & Summers, 1991; Ivancevich, Matteson, Freedman, & Phillips, 1990). The work stress-absenteeism link was proposed as early as 1955 by Hill and Trist (1955). They suggested that absenteeism could be characterized as withdrawal from the work setting due to job stress. They felt that absenteeism was a means that individuals employed in adapting or coping with their environment. Gender as a moderator in stress research, has been suggested (e.g., Steel & Rentsch, 1992) also as an area in need of research. Jick and Mitz (1985) have indicated that intuitively and empirically, some differences between men and women have been observed and one particular emerging area of gender research that they noted is work stress.

Sub-group analysis was performed in this research to investigate the effects of stress in the work setting on employee well-being and absenteeism for males and females within an organizational stress model. Figure 1 depicts the hypothesized path-analytic model guiding this research. The direction of an arrow in Figure 1 indicates the the hypothesized path, while the plus or minus sign indicates the direction of the hypothesized relationship between the two variables. A hypothesis of no relationship between two variables is indicated by the absence of an arrow between them. The exogenous variables in Figure 1 can be divided into two categories: organizational and extraorganizational variables. The organizational variables are depicted in Figure 1 as having direct effects on job stress. The extraorganizational variables in Figure 1 are depicted as directly affecting life stress. The two variables included in this model that have been related in prior research to life stress are family-spouse relationships and financial problems. The endogenous variables in Figure 1 consist of six intervening variables and one outcome variable-absenteeism. These variables were included in the model based on research which



has indicated that job and life stress, emotional exhaustion, depression, cold/flu episodes, somatic symptoms, affect organizational outcomes such as absenteeism. As noted earlier, research has indicated that absenteeism, the single outcome variable used in this research, is not only affected by stress but the stress-absenteeism link may also be moderated by gender (Steel & Rentsch, 1992).

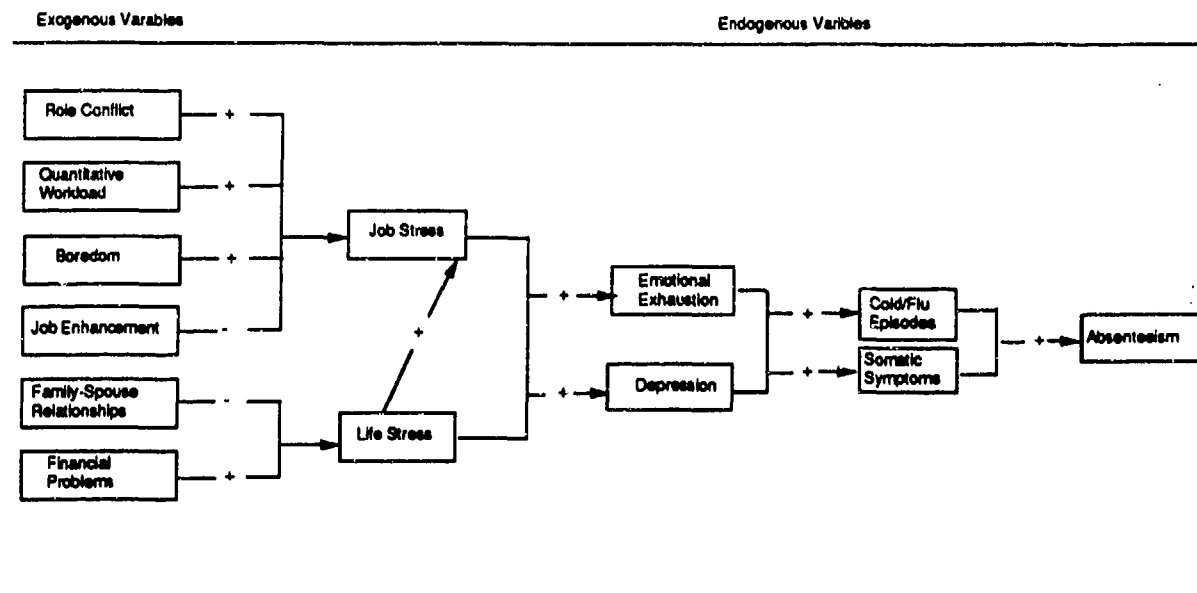


Figure 1. Hypothesized Model

Method

Subjects

Participants were Federal government civilian employees working in government agencies around the country. The sample included 170 males and 204 females. The ages for the participants ranged from 18 to 70 years with an approximate mean of 45 years for males and 42 years for females. Approximately 87% were white, 3% black, 5% Hispanic, and 5% were listed as Indian, Asian, or did not provide their race. Most participants were married (77%) while 9% were single, 12% divorced, and 2% were listed as widows or widowers.

Procedure

At each organizational site a health assessment survey was administered en-masse to the participants. These surveys were completed and collected at the beginning of the seminar before the individuals were exposed to health related information. Absenteeism rates were obtained from the employees' personnel records.

Analysis

Exploratory path analysis was performed separately for males and females using CALIS (SAS, 1990) within the framework presented in Figure 1. CALIS is a

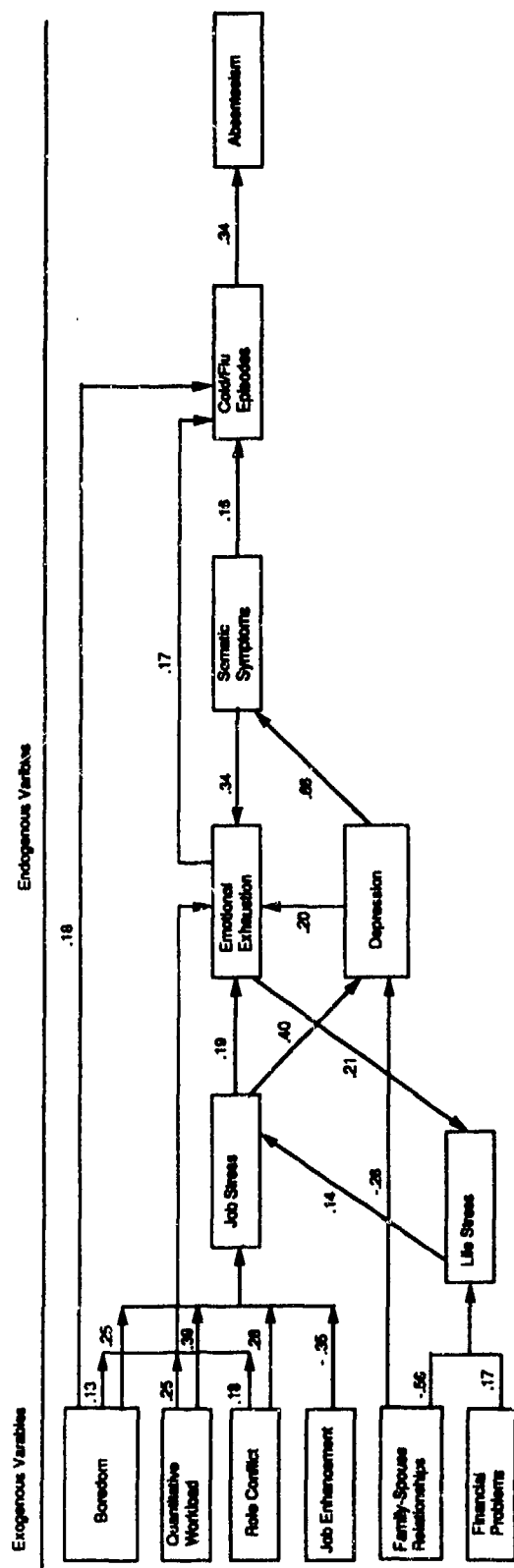


Figure 2. Revised Model For Males

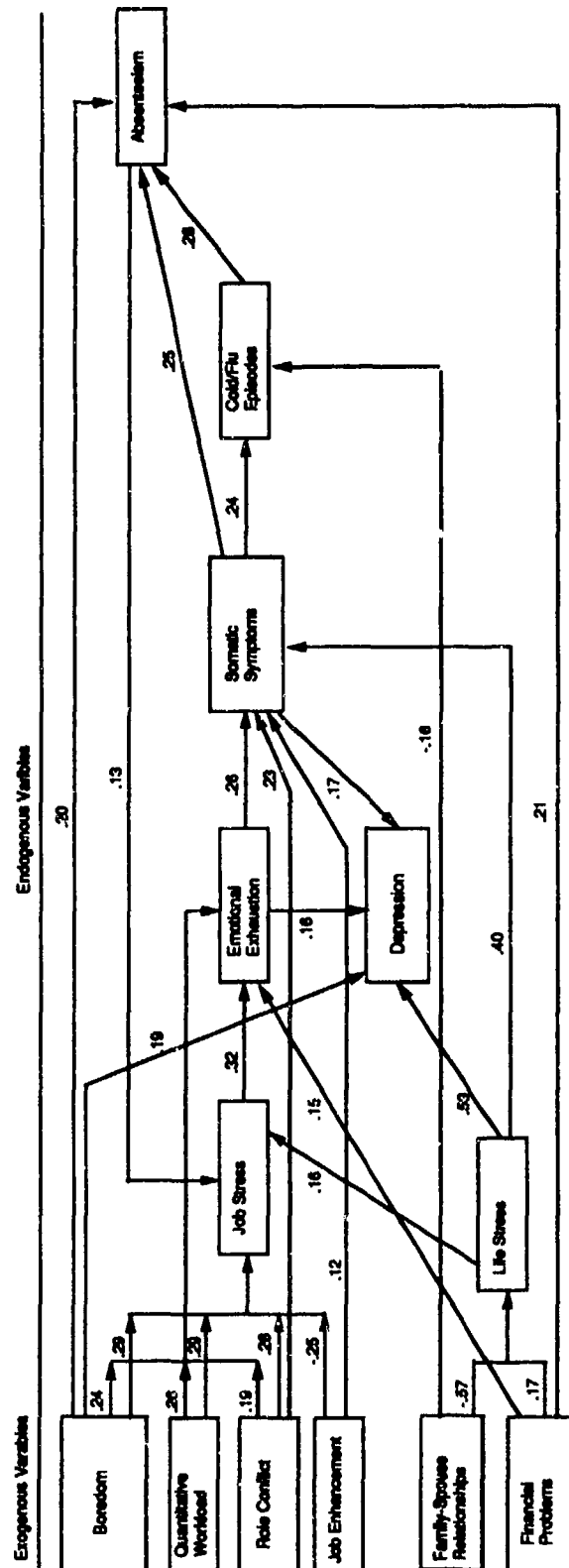


Figure 3. Revised Model For Females

general maximum likelihood structural equation estimation procedure similar to LISREL (Joreskog & Sorbom, 1984) which provides a test of parameter estimates and as assessment of the overall fit of a model. The structural model in figure 1 was represented by two sets of seven linear structural equations, one set for the male model and the remaining set for the female model. The variables found to be non-significant were deleted and the variables that would increase the model fit, as indicated by CALIS's modification indices, were added resulting in the revised models for males (Figure 2) and females (Figure 3).

Results and Conclusions

The path coefficients in Figures 2 and 3 are generally interpreted the same as partial regression coefficients when using regression analysis. In the main, parameter estimates are considered significant if a critical ratio (estimate/standard error) is greater than 2 (Joreskog & Sorbom, 1984, p. III.12). The revised models (Figures 2 and 3) resulting from the path analyses indicated a good fit of the data to both models as indicated by the goodness-of-fit chi-square tests (GFI and AGFI), p values, and root-mean-square values. For the male model the goodness-of-fit values were: $X^2(42) = 45.56$, $p = .33$, GFI = .96, AGFI = .92, and root-mean-square residual = .045. For the females model the goodness-of-fit values were: $X^2(36) = 41.55$, $p = .24$, GFI = .97, AGFI = .93, and root-mean-square residual = .037. In addition, the critical ratios for all paths shown in the revised models were statistically significant (i.e., had values > 2).

Males. Examination of the revised model for males (Figure 2) reveals that 12 of the 17 hypothesized relationships in Figure 1 were supported. The four hypothesized relationships between the exogenous job related variables and job stress were supported. Specifically, boredom, quantitative workload, and role conflict significantly increased job stress while job enhancement decreased it. In turn, the path from job stress to emotional exhaustion, to cold/flu episodes, and finally to absenteeism was supported. The two extraorganizational variables, family-spouse relationships and financial problems, were significantly related to life stress as hypothesized. Good family-spouse relationships reduced life stress while financial problems increased it. In turn, life stress significantly affected job stress but not depression or emotional exhaustion as hypothesized. Depression was directly affected by job stress as hypothesized but also by family-spouse relationships which was not hypothesized to be directly related. Also, somatic symptoms were affected by depression as hypothesized. There were a number of unexpected direct relationships in this model. First, emotional exhaustion directly affected life stress, while somatic symptoms, boredom, quantitative workload, and role conflict affected emotional exhaustion. In addition, boredom directly affected cold/flu episodes while family-spouse relationships directly affected depression.

Females. The revised model for females (Figure 3) is more complicated than hypothesized and more complicated than the revised model for males (Figure 2). The four hypothesized relationships between the exogenous job related variables and job stress were supported as were the relationships between the two extraorganizational variables and life stress. Specifically, boredom, quantitative workload, and role conflict increased job stress while job enhancement decreased it. Also, good family-spouse relationships reduced life stress while financial problems increased it. The

primary endogenous path from job stress to emotional exhaustion, to somatic symptoms, and to absenteeism was supported. Also, the paths from life stress to depression and job stress, and somatic symptoms to absenteeism relationships were supported.

There were more unexpected direct effects for females than for males. First, for females boredom, quantitative workload, role conflict and financial problems all affected emotional exhaustion, while role conflict, job enhancement and life stress affected somatic symptoms. In addition, family-spouse relationships affected the number of cold/flu episodes, somatic symptoms affected depression, absenteeism affected job stress, and lastly, boredom and financial problems directly affected absenteeism for females. An important contrast between the male model and the female model is that absenteeism is predicted by one variable (cold/flu episodes) in the male model and by four (boredom, somatic-symptoms, cold/flu episodes, financial problems) in the female model. Overall, the female model is a more complex model than the male model with 25 significant relationships compared to 21 for the male model.

References

- Ganster, D. C. , & Schaubroeck, J. (1991). Work stress and employee health. Journal of Management, 17(2), 235-271.
- Hendrix, W. H., Steel, R. P., Leap, T. L. & Summers, T. P. (1991). Development of a stress-related health promotion model: Antecedents and organizational effectiveness outcomes. In P. L. Perrewe (Ed.), Handbook on Job Stress (Special Issue). Journal of Social Behavior and Personality, 6(7), 141-162.
- Hill, J. M. & Trist, E. L. (1955). A consideration of industrial accidents as means of withdrawal from the work situation. Human Relations, 8, 121-152.
- Ivancevich, J. M., Matteson, M. T., Freedman, S. A., & Phillips, J. S. (1990). Worksite stress management interventions. American Psychologist, 45(2), 252-261.
- Jick, T. D., & Mitz, L. F. (1985). Sex differences in work stress. Academy of Management Review, 10(3), 408-420.
- Joreskog, K. G., & Sorbom, D. (1984). LISREL VI: Analysis of linear structural relationships by the method of maximum likelihood. 3rd. ed. Mooresville, IN: Scientific Software.
- SAS Technical Report P-200 (1990). SAS/STAT Software: CALIS and LOGISTIC procedures, release 6.04. Cary, NC: SAS Institute.
- Steel, R. P., & Rentsch, J. R. (1992). Proximal and distal predictors of absence from work. Unpublished manuscript.



Determinants of Selective Processing of Emotion Cues

Maryanne Martin, D.Phil.

University of Oxford

Gregory V. Jones, Ph.D.

University of Warwick

92-17688



Abstract

Are there systematic differences among people in the relation between their cognitive and emotional processes? An experiment is reported in which the effects of gender and of trait anxiety upon selective processing of threatening stimuli are investigated. Data from 111 participants in a computerised Stroop task are reported. The effects of gender and trait anxiety were found to interact with the nature of the task. For men, there was little difference between the processing of either threatening or positive stimuli by either high-anxiety or low-anxiety individuals. For women, on the other hand, reaction times were greater for threatening than for positive stimuli for high-anxiety individuals, but were greater for positive than for threatening stimuli for low-anxiety individuals. The implications of these findings are discussed.

In recent years, considerable progress has been made in analysing the links between emotional and cognitive processes. In particular, it has been established that both the perception and remembering of emotional material is influenced by the current emotional state of the person concerned (e.g., see Eysenck, in press; Williams, Watts, MacLeod & Mathews, 1988). For example, it has been shown that a mood of anxiety can lead to selective processing of threat cues (Martin, Williams, & Clark, 1991; Mathews & MacLeod, 1985). At the most general level, there is reason to believe that these cognitive effects are an integral feature of a person's mental architecture, rather than being entirely the product of learning (Jones & Martin, 1992; Martin, Horder, & Jones, in press; Oatley & Johnson-Laird, 1987). As yet, however, there has been relatively little investigation of the extent to which the cognitive effects of emotion might differ between different individuals.

The present work is directed toward exploring a particular instance of how the cognitive effects of emotion might vary among individuals. The task is a test of selective attention using materials which are either threatening, neutral or positive. Since materials which are threatening act as a cue for anxiety, it is possible that the processing of threatening materials differs systematically among individuals as a function of the individuals' normal levels of anxiety. Most work in this area has compared anxious patients with the normal population. There has been little examination of possible effects of different levels of anxiety within the normal population. It should be particularly informative to examine whether such effects can be observed even among a highly selected and generally healthy population.

A further possibility is that the processing of threatening materials is also a function of a person's gender, and indeed that the two characteristics of trait

anxiety and gender might be interactive in their influence upon the cognitive processing of threatening material. In view of the greater incidence of women patients among those suffering from emotional disorders, there has been a tendency for work in the area of cognition and anxiety to focus upon female participants. There is thus at present a scarcity of experimental studies with sizeable groups of both male and female participants. In the experiment to be reported, selective processing of threatening material was examined as a function of trait anxiety in both women and men.

Method

Subjects

There were 111 participants in the experiment (data from four further people were incomplete and thus they were not included). All were university undergraduates, with mean age 20.8 (SD 2.8) years. There were 49 females and 62 males.

Apparatus

All instructions and stimuli were displayed on a colour monitor controlled by a Ness AT computer. Latencies of spoken responses were logged automatically into this computer via a voice key with microphone.

Materials

Threatening, neutral, and positive word lists were compiled. There were 24 threatening words. These were drawn from those used by Mathews and MacLeod (1985, 1986) and MacLeod, Mathews, and Tata (1986). Half were physically threatening (e.g., "injured") and half were socially threatening (e.g., "indecisive"). There were 12 neutral words (e.g., "occasional") and 12 positive words (e.g., "peaceful"). These were drawn from those used in Experiment 4 of Martin, Williams, and Clark (1991). The threatening words and the positive words were matched in terms of rated emotionality, but were of opposite polarities. Each word was presented once in each of four different colors. These were blue, green, pink, and red.

Procedure

Participants were tested in cubicles on a series of tasks. The trait anxiety of each participant was measured using Spielberger's state/trait anxiety scale. Following this, the participant performed a Stroop task (see Dyer, 1973) presented via a colour monitor. On each trial, a row of 13 white asterisks first appeared in the centre of the screen for 500 msec to assist fixation (the longest stimulus word was 11 letters). It was followed immediately by a colored stimulus word. The participant had to name the colour in which the word appeared as quickly as possible. The word remained on the screen until the participant's response was recorded by a voice key. The monitor then remained blank for 1 sec until the next trial commenced. The order of presentation of stimuli was randomized across trials. The experimental session was preceded by a practice session using an additional set of 12 neutral words, for which data were not collected.

Results

Mean levels of trait anxiety were 43.3 and 42.5 for women and men, respectively, and did not differ significantly, $t(109) = 0.39$. Participants were divided at the median trait-anxiety score into a low anxiety (≤ 41) group and a high anxiety (≥ 42) group. There were 24 low-anxiety and 25 high-anxiety females, with 28 low-anxiety and 34 high-anxiety males. Median latencies were calculated for each of the three conditions for each participant. The means of these values for the four groups are shown in Table 1.

Table 1

Mean Latencies (msec) as a Function of Gender and Trait Anxiety

Gender	Anxiety	Word type		
		Positive	Neutral	Threat
Male	Low	649	647	652
Male	High	626	633	628
Female	Low	732	709	712
Female	High	605	628	629

The data summarised in Table 1 were subjected to an analysis of variance with two between-subjects variables (gender and trait-anxiety level) and one within-subjects variable (word type). There was a significant two-way interaction between anxiety level and word type, $F(2, 214) = 8.01$, $p < 0.001$. This interaction was itself modulated by a significant three-way interaction between gender, anxiety level, and word type, $F(2, 214) = 5.35$, $p < 0.01$. It can be seen from Table 1 that for high-anxiety women the mean latency for threatening stimuli was greater than that for positive stimuli, whereas for low-anxiety women the pattern was reversed. In contrast, for both high-anxiety and low-anxiety men there was little difference between mean latencies for threatening and positive stimuli. No other effects reached significance, although the main effect of trait anxiety almost did so.

Discussion

The results of the present experiment show that even within the normal population individuals differ significantly in how their cognitive processes are affected by threatening stimuli. Women with high levels of trait anxiety were more distracted by threat cues than by positive cues, whereas the converse held for women with low levels of trait anxiety. This finding is comparable to that arising when anxious patients are compared with normal control groups. Anxious patients have been found to be relatively distracted by threatening material, and control participants by nonthreatening material (Martin, Williams, & Clark, 1991, Experiment 2; Mathews & MacLeod, 1985).

The greater incidence of women in the clinical population suffering from emotional disorders suggests the possibility that women may be relatively sensitive to emotional material. Similarly, Clark and Teasdale (1985) reported for the normal population a mood congruency effect occurring with women but not with men. Women recalled more positive than negative words in a happy mood, and more negative than positive words in a depressed mood. This finding is analogous to the present one if it is borne in mind that in one case emotional sensitivity is manifested as greater interference in naming, while in the other it is manifested as higher recall. Why might a gender difference arise in the cognitive effects of emotion even when, as here, there was no gender difference in the overall level of emotion? According to a propagational view of emotion (Martin & Williams, 1990), cognitive effects of emotion tend to act to maintain or enhance that emotion. In some situations such maintenance is likely to be advantageous, and in others it is likely to be disadvantageous. It may be that a gender difference has arisen as a consequence of differential exposure to the two types of situation.

References

- Clark, D. M., & Teasdale, J. D. (1985). Constraints on the effects of mood on memory. Journal of Personality and Social Psychology, 48, 1595-1608.
- Dyer, F. N. (1973). The Stroop phenomenon and its use in the study of perceptual, cognitive and response processes. Memory & Cognition, 1, 106-120.
- Eysenck, M. W. (in press). Anxiety: The cognitive perspective. London: Lawrence Erlbaum Associates.
- Jones, G. V., & Martin, M. (1992). Conjunction in the language of emotions. Manuscript submitted for publication.
- MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. Journal of Abnormal Psychology, 95, 15-20.
- Martin, M., Horder, P., & Jones, G. V. (in press). Integral bias in naming of phobia-related words. Cognition and Emotion.
- Martin, M., & Williams, R. M. (1990). Imagery and emotion: Clinical and experimental approaches. In P. J. Hampson, D. F. Marks, & J. T. E. Richardson (Eds.), Imagery: Current developments (pp. 268-306). London: Routledge.
- 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.
- Mathews, A. M., & MacLeod, C. (1985). Selective processing of threat cues in anxiety states. Behaviour Research and Therapy, 23, 563-569.
- Mathews, A. M., & MacLeod, C. (1986). Discrimination of threat cues without awareness in anxiety states. Journal of Abnormal Psychology, 95, 131-138.

Oatley, K., & Johnson-Laird, P. N. (1987). Towards a cognitive theory of emotions. Cognition and Emotion, 1, 29-50.

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



Women in the Military: Comparison of Attitudes and Knowledge of
Service Academy Cadets Versus Private College Students¹

Michael D. Matthews
Drury College

Abstract

A survey was administered to service academy cadets and private college students to assess their attitudes and knowledge concerning the role of women in the military. Results indicated that service academy cadets, particularly males, were less approving of women serving in a variety of military roles than were private college students. This was especially evident for combat roles. With respect to knowledge, service academy cadets were more knowledgeable than private college students concerning the current roles women play in the military.

American's attitudes toward the role of women in the military have changed substantially since World War II. The widely publicized role of women in recent American military ventures such as the Panama invasion and Project Desert Storm underscore the increasingly active part women play in armed conflict, and may serve as a catalyst for further changes in official military policy toward the participation of women in combat settings.

Public attitudes toward the role of women in the military may influence the evolution of military policy in this domain. Recent studies suggest that, while Americans are in favor of a more complete integration of women into the military, they are not for total equality between the sexes (Davis, Lauby, & Sheatsley, 1983). Moreover, Matthews, Melton, & Weaver (1988) provided evidence that women are generally more supportive of increased roles for women in the military than are men, and blacks are less supportive than whites. Additionally, Matthews and Weaver (1990) reported that several other demographic variables affect attitudes in this domain. In general, younger, more highly educated persons employed in white collar jobs who perceive themselves to be relatively well-off financially are more likely to support expanded roles for women in the military.

Because education is significantly related to attitudes toward women in the military, it seems logical that the type and setting in which higher education is experienced might also relate to attitudes and knowledge about women's roles in the military. At one extreme, one might consider the attitudes of young men and women who elect to attend a service academy. In addition to being socialized into the military, persons with certain attitudes might be attracted to a military setting. It is commonly maintained that the service academies attract more conservative students than other colleges and universities. This is somewhat paradoxical when considering women's roles in the military, because many current jobs that women fill in the military might be defined as "nontraditional." Thus, one might expect some role conflict, especially among women cadets, when placed into an environment which encourages behaviors and activities that conservatives might feel are inappropriate for women in our society. The education, indoctrinization, and experiences of women cadets would presumably serve to reduce this conflict.

In contrast, students who elect to attend a traditional college or university may be somewhat more liberal than service academy cadets and probably receive less information during their education about the military in general and women's roles in the military in particular.

The purpose of the current study was to compare the attitudes and knowledge of service academy cadets and students from a private liberal arts college with respect to women's roles in the military. Because of the previously noted differences of attitudes of men and women toward this issue (Matthews et al., 1988), sex of respondent was also included as a variable. For male respondents,

¹ The author expresses his sincere appreciation to Major Gordon Curphy and Major Karen Dunivin of the USAFA for assistance in data collection and reduction

one might expect different attitudes for students from the two institutions. Although it has been demonstrated that college provides a "liberalizing" influence (e.g., Newcomb, 1943), one might not see this as much among the male service academy cadets, due to selection bias and the importance of tradition and conformity among military cadets. Indeed, there is considerable anecdotal evidence concerning the resistance of male cadets toward sexual integration of the academies. Thus, one might predict that male students at the private liberal arts college would express more liberal attitudes toward women's roles in the military than male cadets. With female respondents, predictions are less clear. Female cadets, in fact, may have better formed opinions concerning this issue than their private college counterparts. Both men and women cadets should also have more accurate knowledge concerning actual roles of women in today's military than private college students.

Method

Students enrolled in introductory psychology classes at a United States military service academy (N = 393 males; 67 females) and a private liberal arts college located in the midwest (N = 25 males; 52 females) served as subjects. At the service academy cadets ordinarily enroll in the beginning psychology class during their first year of study. The majority of students at the private college were freshman, but approximately 30 percent were upperclassmen, mostly sophomores. Thus, the majority of respondents from both institutions were between 18 and 20 years of age, and represented numerous academic majors.

The respondents were asked two sets of questions. The first required them to indicate whether a woman "should" or "should not" be allowed to serve in the following military roles: nurse in combat zone, typist in the Pentagon, commander of a military installation, truck mechanic, jet transport pilot, jet fighter pilot, anti-aircraft gunner in the United States, hand-to-hand combat soldier, and crew member on a combat ship. These jobs provide a mix of traditional and nontraditional roles for women in the military.

The second set of questions focused on knowledge about women's roles in the military. They were worded as follows:

As far as you know, are women now assigned to jobs in the armed forces that would expose them to combat, or are women not assigned to such jobs? (Yes, they are; no, they are not; don't know) [ASSIGNCOMB]

As far as you know, are women in the military now assigned to dirty jobs like repairing trucks or other heavy equipment, or are women not assigned to such jobs? (Yes, they are; no, they are not, don't know) [ASSIGNDIRT]

As far as you know, are women in the armed services now assigned to jobs where they have command over men, or are women not assigned to such jobs? (Yes, they are; no, they are not; don't know) [ASSIGNMEN]

Finally, one additional question was asked that sought the respondent's perception of the opportunities for women in the military versus the civilian environment. It read as follows:

Thinking about opportunities and equal treatment for women, would you say their treatment and opportunities are better in the military, better in civilian employment, or that there isn't any difference these days? (Better in the military; better in civilian employment; no difference) [WOMENOPP]

These questions were taken from the General Social Surveys (see Davis & Smith, 1990).

It might be noted that the data were collected following Operation Desert Storm. The publicity surrounding this military action may have had some impact on responses, especially among the civilian sample.

Table 1

Percentage of Respondents Indicating Women Should Serve in Specified Military Roles As a Function of Institution and Sex

Role	Service Academy		Private College	
	Men	Women	Men	Women
Fighter Pilot	42.2%	39.6%	60.0%	94.2%
Truck Mechanic	86.8%	97.0%	88.0%	94.2%
Nurse in Combat Zone	89.1%	100.0%	92.0%	96.2%
Typist in Pentagon	97.2%	95.5%	100.0%	98.1%
Military Commander	62.3%	94.0%	48.0%	73.1%
Hand-to-Hand Combat Soldier	21.6%	56.7%	32.0%	57.7%
Jet Transport Pilot	86.8%	98.5%	88.0%	98.1%
Air Defense Gunner	76.6%	89.6%	56.0%	80.8%
Crew Member or Combat Ship	58.5%	86.6%	80.0%	88.5%
	(N=393)	(N=67)	(N=25)	(N=52)

Results and Discussion

Table 1 compares the percentage of respondents indicating women "should" serve in the roles defined above as a function of institution and sex. Note that, in general, service academy cadets were less favorable toward women's full participation in the military than were students from the private college. Also, there is a reliable pattern for women respondents from both institutions to hold more favorable views toward expanded roles for women.

The responses to the knowledge questions and the question dealing with military versus civilian opportunities for women, as a function of institution and sex, are summarized in Table 2. On the three knowledge questions cadets, as might be expected, respond more accurately. Note also the relatively large percentage of "Don't Know" responses among the civilian respondents. Sex differences in responses to these questions are not as pronounced nor as consistent as among the attitudes questions. On the question dealing with opportunities and treatment of women, cadets of both sexes expressed that the military presents better possibilities, while the private college students perceived better chances in the civilian world.

Several trends emerge from these data. First, service academy cadets are much more knowledgeable about military roles that women currently play than are students at the private liberal arts college. Second, service academy cadets, especially men, are much more conservative in their attitudes concerning the role of women in the military than are their civilian college counterparts. Finally, consistent with observations from the general population noted by Matthews et al., (1988), women from both institutions are more liberal in their attitudes than the men. In general, the disparities between the men and women and between the service academy cadets and the private college students were greatest in nontraditional combat roles. This is also consistent with earlier observations

by Matthews et al. (1988) and Matthews and Weaver (1990).

A discussion of why service academy cadets tend to respond somewhat conservatively is beyond the scope of the current discussion. However, selection bias and the general conservative nature of the military may play a role. For male cadets, the long-standing tradition of the military as a bastion of masculinity may also contribute to their relatively conservative responses.

Table 2
Responses to Knowledge and Opportunity Questions
by Institution and Sex

QUESTION	Service Academy		Private College	
	MALE	FEMALE	MALE	FEMALE
ASSIGNCOMB				
YES	70.5%	80.6%	48.0%	48.1%
NO	24.7%	16.4%	20.0%	23.1%
DK	4.6%	3.0%	32.0%	28.8%
ASSIGNDIRT				
YES	79.6%	83.6%	60.0%	44.2%
NO	5.1%	3.0%	8.0%	5.8%
DK	15.0%	13.4%	32.0%	50.0%
ASSIGNMEN				
YES	96.4%	98.5%	64.0%	63.5%
NO	.5%	---	4.0%	7.7%
DK	2.5%	1.5%	32.0%	28.8%
WOMENOPP				
MIL BETTER	40.5%	37.3%	---	15.4%
CIV BETTER	33.1%	35.8%	80.0%	57.7%
NO DIFF	25.4%	26.9%	20.0%	26.9%
	(N=393)	(N=67)	(N=25)	(N=52)

References

- Davis, J. A., & Smith, T. W. (1990). General social surveys, 1972-1990: Cumulative codebook. Chicago: National Opinion Research Center.
- Davis, J.A., Lauby, J., & Sheatsley, P.B. (1983). American's view the military. (Tech Report/NORC Rep. No. 131). Chicago: National Opinion Research Center.

- Matthews, M.D., Melton, E.C., & Weaver, C.N. (1988). Attitudes toward women's roles in the military as a function of gender and race. Proceedings of Psychology in the Department of Defense, Eleventh Symposium. Colorado Springs, CO: U.S. Air Force Academy.
- Matthews, M. D., & Weaver, C. N. (1990). Demographic and attitudinal correlates of women's role in the military. Proceedings of Psychology in the Department of Defense, Twelfth Symposium. (USAF TR-90-1). Colorado Springs, CO: U.S. Air Force Academy.
- Newcomb, T. M. (1943) Personality and social change. Attitude formation in a student community. New York: Dryden Press.



Gender and Performance in Naval Aviation Training

Annette G. Baisden
Naval Aerospace Medical Institute

Abstract

Women have played an important role in naval aviation since their entry into naval aviation training as pilots in 1973 and as naval flight officers in 1979. Heightened by the Congressional debates over women in combat, questions are being raised regarding equitable selection, fair treatment in training, and equality of opportunity. This paper presents a brief historical perspective of the rationale and research related to women in military aviation, examines selection and training data of 13,755 men and 421 women who entered naval aviation training from 1984 to 1991, and addresses the issues of equitable selection and fairness in naval aviation training. Analyses of the data include descriptive statistics and t tests to compare male/female performance on selection tests and pre-flight training grades, a test of equal proportions and a chi-square test to assess differences in attrition, correlation and regression analyses to determine the significance of relationships between selection test measures and performance in pre-flight training. Analysis of gender differences indicated that women had significantly better scores than men on the aviation selection tests which are predictive of pre-flight academic training performance ($p < .01$). Their performance grades during the pre-flight academic training, however, were significantly lower than that of men ($p < .01$). Attrition rates and types of attrition did not differ. Implications of the findings and future directions for research are suggested.

Women have played an important role in naval aviation since their entry into naval aviation training as pilots in 1973 and as naval flight officers in 1979. Their entry was due to the lifting of the number restrictions imposed by the Women's Armed Services Act of 1948, the manpower gap created by the all volunteer force, and the impact of the women's liberation movement. Heightened by the current congressional debate regarding the role of women in combat, questions are being asked regarding equitable selection, fair treatment in training, and equality in opportunities. This paper presents a brief historical perspective of research related to women in military aviation, examines recent selection and training data, and addresses issues of equitable selection and fairness in naval aviation training.

By law, women pilots and naval flight officers are precluded from permanent assignment to squadrons with combat missions. Currently, women comprise two percent of the naval aviation officer population. The limitations on squadron assignments and the lack of integrated fleet weapons and tactical experience promote differences in the career development and satisfaction of men and women. Major problem areas with regard to the assimilation of women are their acceptance, their utilization, and the quality with which they are treated (1990 Navy Women's Study Group, 1991). The Navy's task is to determine what can be done to facilitate the assimilation of women.

92-17690



Early research focused on finding solutions to problems encountered in the introduction of the first women into naval aviation training. Willingness of men to accept the new role of women in one of the most traditional of all service groups was the impetus of a survey of attitudes by Baisden (1974). Peer attitudes were generally accepting; however, senior officers felt women were out of place in a combat-oriented organization. Concern about individual differences resulted in a pilot personality profile that transcended sex distinctions (Novello & Youssef, 1974). The question of grounding related to menstrual cycle and to oral contraceptives prompted the respective studies of Baisden and Gibson (1975) and Altekruze and Sanders (1976). Studies addressing difficulties in physical training and survival training led to new aerobic conditioning requirements and survival techniques that compensated for female deficiencies in upper-body strength and endurance (Surburg & Williamson, 1974). Requirements to ascertain the strength capabilities needed to control aircraft under abnormal conditions led to the definition of aircraft control forces and female strength characteristics and measuring techniques for screening pilots (McDaniel, 1978). Women's sizes caused reevaluation of standards, aircraft restrictions, and aircraft designs relative to anthropometric measures (Ketchan-Wiedl & Bitner, 1976).

As the number of women in aviation training increased, research shifted to performance evaluation. Kantor, Nobel, Leisey, and McFarlane (1979) and Baker (1989) found comparable performance of Air Force men and women on selection measures, training measures and completion rates. Morvant (1981), however, found the flying performance of Air Force women significantly inferior to that of men and raised the issue of male bias and objectivity of flight grades. Carretta (1990) found the graduation rate from Air Force training was significantly higher for men and attributed the difference to women's lower selection scores. His findings were consistent with those of Siem and Sawin (1990) evaluating Air Force women and with Brown and Dohme (1980) evaluating Army students in aviation training.

The objectives of this study are to determine whether there are gender differences in 1) selection test scores of students entering naval aviation training, 2) attritions rates and types of attrition, and 3) grades in pre-flight training.

Method

Selection and training data for 13,755 men and 421 women who entered naval aviation training to become aviators or naval flight officers between October 1984 and December 1991 were examined. These numbers represent the total population (excluding Marines) of students. Selection data included scores on the Navy's Aviation Selection Test Battery, paper-and-pencil tests consisting of the Academic Qualifications Test (AQT) and the Flight Aptitude Rating (FAR). Training data included academic grades, peer ratings, instructor evaluations, final grades during pre-flight training, and attrition statistics.

Analyses of the data included descriptive statistics and *t* tests for comparison of performance on selection tests and training grades, a test of equal proportions and a chi-square test for assessing differences in attrition, correlations and regressions to determine significance of relationships between selection test measures and performance in pre-flight training, and regression analysis to examine test bias.

Results and Conclusion

Gender differences in selection test scores of aviation students were highly significant. Women achieved higher scores on the AQT ($t(14172) = -5.83, p < .001$), but men achieved higher scores on the FAR ($t(14159) = 19.01, p < .001$). These findings are consistent with those reported by Blower, Dolgin and Schull (1991) with a sample of applicants for naval aviation training.

The correlations of AQT and final grade in pre-flight training for women ($r = .30, p < .001$), and for men ($r = .36, p < .001$) were highly significant and support the historical trend reported by Baisden and Holcombe (1991).

Because women had higher AQT scores, a higher completion rate in pre-flight training was expected. A test for equal proportions indicated the 17% attrition rate for men was not significantly different from the 18.4% attrition rate for women ($z = .61$). Differences in types of attrition were not significant ($X^2(3, N = 2040) = .863, n.s.$).

The relationship between gender and performance in training was addressed by examining test scores with gender as predictor variables. A regression analysis of the complete/attrite variable on AQT score and gender was performed. An SPSS stepwise procedure using default coefficient significance test produced the equation: $\text{Complete} = .660 + .033(\text{AQT})$. Gender entered the equation but was removed indicating lack of predictive significance. When final grade was used as the dependent variable, gender remained in the equation. $\text{Final} = 43.40 + 1.37(\text{AQT}) - 2.95(\text{sex})$.

Results of t tests comparing the means of pre-flight training grades indicated the grades of men were significantly higher than those of women ($p < .001$). Kirkland (1978) reported that females are more reactive to criticism during the learning process. This reactivity, college majors, disposition toward peer support, and systematic differences in both acceptance and equality of treatment may explain some of the performance differences and, therefore, warrant further study.

References

- Altekruse, E. B., & Sanders, M. G. (1976, May). Drugs in aviation medicine: An investigation of the effects of oral contraceptives on performance. Paper presented at the meeting of the Aerospace Medical Association, Miami, FL.
- Baisden, A. G. (1974, September). Attitudes of military personnel toward women with specific reference to naval aviation. In R. K. Ambler (Chair), Women in aviation: Progress and problems. Symposium conducted at the meeting of the 82d Annual Convention of the American Psychological Association, New Orleans, LA.
- Baisden, A. G., & Gibson, R. S. (1975, October). Effects of the menstrual cycle on the performance of complex-perceptive motor tasks. Proceedings of the Human Factors Society 19th Annual Meeting, 415-417.
- Baisden, A.G., & Holcombe, F.D. (1991). Aerospace psychological qualifications. In U.S. Naval Flight Surgeons Manual (pp. 1-5). Washington, D.C.: Bureau of Medicine and Surgery.

- Baker, L. E. (1989). The effect of higher education variables on cadet performance during 1987 light aircraft training (AU-ARI-88-9). Maxwell AFB, AL: Air University.
- Blower, D. J., Dolgin, D. L., & Schull, R. N. (1990). Naval aviation selection test scores and female aviation performance. Proceedings of the AGARD Symposium on Recruiting, Selection, Training, and Military Operations of Female Aircrew, (pp 6.1-6.5). Tours, France.
- Brown, W. R., & Dohme, J. A. (1980). An evaluation of minority and female performance in Army rotary wing aviation training: Volume II Evaluation report (Research Report 1319). Fort Rucker, AL: US Army Research Institute for the Behavioral and Social Sciences.
- Carretta, T. R. (1990, April) Gender differences in USAF pilot training performance. Paper presented at the meeting of the 12th Psychology in the DOD Symposium, Colorado Springs, CO.
- Kantor, J. E., Noble, B. E., & Leisey, S. A., & McFarlane, T. (1979). Air Force female pilots program: Initial performance and attitudes (AFHRL-TR-78-87). Brooks AFB, TX: Air Force Human Resources Laboratory.
- Ketchan-Wiedl, M. A., & Bitner, A. C. (1976). Anthropometric accomodation of a female population in a workplace designed to male standards (PMTTC-TP-76-3). Pt. Mugu, CA: Guided Missile Test Center.
- Kirkland, F. (1978). Utilization of women in the Navy. Minutes of the Tenth Training and Personnel Technology Conference. Washington, D.C.: Office of the Under Secretary of Defense.
- McDaniel, J.W. (1978). A strength screening program for the USAF pilot candidates. Minutes of the Tenth Training and Personal Technology Conference. Washington, D.C.: Office of the Under Secretary of Defense.
- Morvant, D.A., & Thrower, M.B. (1981). An Analysis of the Training and Performance of Women in Undergraduate Pilot Training (Research Report 1319). Maxwell AFB, AL: Air Command and Staff College, Air University.
- 1990 Navy Women's Study Group. (1991). An update report on the progress of women in the Navy. Washington, D. C.: Department of the Navy.
- Novello, J. R., & Youssef, Z. I. (1974). Psycho-social studies in general aviation II. Personality profile of female aviators. Aerospace Medicine, 45, 630-633.
- Siem, F. M., & Sawin, L.L. (1990). Comparison of male and female USAF pilot candidates. Proceedings of the AGARD Symposium on Recruiting, Selection, Training, and Military Operations of Female Aircrew (pp.5.1-5.4). Tours, France.
- Surburg, P. R., & Williamson, C. B. (1974) A study to establish physical fitness norms for female aviators. Pensacola, FL: University of West Florida.



Mental Health Issues:
Lessons From Operation Desert Shield/Storm

James A. Martin, Ph.D.
Colonel, U.S. Army
Commander, U.S. Army Medical Research Unit - Europe

Abstract

This paper highlights mental health issues from Operation Desert Shield/Storm (ODS). It is based on information reported at an ODS Lessons Learned Conference held at Vilseck, Germany in October 1991.

During the period 1 August 1990 to 1 August 1991 there were only 476 psychiatric evacuations to U.S. Army medical facilities in Europe from the 500,000 member force deployed to Southwest Asia. (This represents 6.5% of all evacuations during this time period and Europe was the primary evacuation route for all of Southwest Asia). Together with the very low incident rates of "problem behavior" and the tremendously successful air and ground campaigns, it is clear that America deployed a psychologically healthy military force, including combat units that were well trained and well led.

This war was not the challenge it could have been (Knudson, 1991). Yet, there were still service members exposed to the extreme stresses of war (inflicting and/or observing death, inflicting or being the victim of friendly fire, experiencing the horror of human remains, and encountering the universal victims of war - the men, women and children killed, wounded, or made homeless). These service members represent "pockets of trauma," individuals who are potential post-war mental health casualties.

From a mental health perspective, the primary lesson in ODS centers on inadequate Army preparations to meet the mental health needs of its deployed force (before, during, or after combat). Unlike active component combat units, many active and reserve Army mental health units (and personnel) deployed to ODS were not adequately prepared. They were not well equipped. In many cases mental health doctrine was not followed. The Army's reliance on reserve component mental health units was also not successful.

Basis of Assessment

This information comes primarily from material presented by participants at a combined Combat Psychiatry Training Course and

The views expressed in this document do not purport to reflect the position of the Department of the Army or the Department of Defense (PARA 4-3, AR 360-5).

92-17691



Operation Desert Shield/Storm Lessons Learned Conference co-sponsored by Headquarters, 7th Medical Command and the Walter Reed Army Institute of Research. This conference took place at the U.S. Army Combat Training Center in Vilseck, Germany from 6 through 9 October 1991. Twenty-eight of the participants were mental health officers selected because of their experiences during Operation Desert Shield/Storm. Ten of these served at corps or division level assignments during the ground war. They included both active and reserve component mental health officers. In addition, four of these individuals (including the author) participated in earlier mental health after-action lessons learned meetings conducted by the Army Surgeon General's Office and the Army Medical Department Center and School.

Major Lessons Learned and Recommendations

Overall Preparation

The overall organization, staffing, equipment, and training of mental health units deployed for ODS were severely inadequate.

No provisions were made prior to deployment for the mental health needs of armored cavalry regiments or separate brigades. These units were eventually task-organized so that they became the size of peacetime divisions. Two of these units suffered the majority of all U.S. casualties during the ground fighting.

With a few notable exceptions, division mental health teams were not effective in providing consultation to commanders or direct psychiatric services to soldiers. Many division mental health officers and enlisted personnel were too junior (in grade and experience) to elicit the respect and confidence of mid to senior level commanders. Many lacked basic soldier skills. Military experience, including the ability to identify with the line, was even more critical than rank as a factor in successful consultation. At the time of deployment notification, many of the division mental health teams lacked critical officer and enlisted personnel. In the two divisions deployed from Germany, four of six officers and a corresponding number of enlisted specialists had to be designated at the last minute because positions were vacant or individuals were non-deployable. Overall, five of the seven Army divisions deployed to ODS without their full complement of personnel. Some filler personnel were not adequately trained and/or qualified for their positions. Again, with the exception of two divisions, these mental health teams had not participated in any significant field training exercises. Many did not even have or did not know the location of their TO&E unit equipment (especially vehicles) prior to notification.

Even before the deployment, the Army's Medical Detachment Psychiatric (OM) Team concept was known to be obsolete and deficient in mobility and patient holding capability. These 50 person teams support corps and theater level mental health requirements. The Combat Stress Control Companies and Combat

Stress Control Detachments (CSCCs & CSCDs) proposed in the Army's Medical Force 2000 (FM 8-51, 1991) were in the final stage of approval, but the OM teams were not allowed to upgrade to the proposed standard. Only three OM teams were deployed to Southwest Asia (all at "C3-minus" status - missing key personnel and/or equipment) and they either never received their full allotment of personnel or if they did, these personnel did not arrive in a timely manner. Two teams had to replace their commanders before deployment. The three teams deployed represented only 40% of the minimum basis of allocation by current recognized planning standards. With their deficiencies in TO&E and fielding, they represented about 25% of the minimum requirement for mental health capability expected in ODS. Deployed OM Teams were not adequately staffed, equipped, or trained. Had we faced the worst case scenario for surgical casualties, the medical evacuation system might have been severely stressed by inappropriate psychiatric evacuations. Even with the best case scenario which occurred, a direct consequence of late deployment of OM Teams was excessive evacuation of soldiers with minor psychiatric problems and psychosomatic conditions during the early phase of ODS.

Based on these experiences, the TO&Es of armored cavalry units and separate brigades should include some minimal mental health staffing. Division mental health positions need to be filled with more senior professionals and these teams must be given the opportunity to regularly train (and operate) in divisional training exercises. (The most effective ODS combat stress control efforts focused at brigade level and this should be the model for peacetime as well as wartime combat stress control operations). Finally the Army's proposed Medical Force 2000 (MF2K) combat stress control companies and detachments should be fielded from existing resources as soon as possible. These units should operate according to doctrine, including their use in response to disaster and other humanitarian missions.

Stress Debriefings

After Action Stress Debriefings (similar to what the civilian literature labels Critical Incident Stress Debriefings) are a recognized method of providing "first aid" and/or "a preventive inoculation" to individuals who have been exposed to extreme trauma (Shalev & Ursano, 1990). Debriefings are thought to reduce the incidence of Post Traumatic Stress Disorder. The Army also considers stress debriefings important in preventing battle fatigue (FM 8-51, 1991).

Even in a war that was extremely successful, very brief, and with few American casualties, as previously described, there were recognizable "pockets of trauma." The debriefings performed with small units on the battlefield legitimized the expression of feelings, helped individuals understand that their emotions and corresponding stress symptoms were normal reactions, and provided individuals an opportunity to gain some cognitive mastery over

the stressful event(s). These debriefings were also thought to enhance group cohesion.

Unfortunately, After Action Stress Debriefings were not consistently performed during the post-war period. Some mental health personnel/units did not recognize this as part of their mission and others were not adequately trained to perform this service. The situation was worsened by the general lack of mental health consultation to commanders before combat. The lack of peacetime consultation to unit leaders created a situation where leaders were not always receptive to mental health support. When debriefings were performed, soldiers were very appreciative and there were often noticeable positive changes in individual and small group behaviors and attitudes.

After Action Stress Debriefings were important therapeutic interventions for ODS units (and soldiers) exposed to extreme stress (Martin, 1991). Doctrine should require mental health intervention (stress debriefings) whenever a unit experiences extreme trauma (fatal training or operational accidents, terrorism, or combat exposure) and mental health personnel must be prepared for this mission.

Focus on Return to Duty

ODS clearly demonstrated that combat leaders and senior medical officers usually desire immediate evacuation of all casualties out of the theater of operations. Leaders are not focused on returning the soldier to his unit so that he can fight in the next battle. The line commander wants to get his injured soldier to the best possible medical care, and to get him "out of the way of any additional danger." The line commander is focused on a quick, decisive battle, starting with sufficient forces so that replacements are not an issue.

During ODS senior medical officers were focused on surgical wound casualties whose potential arrival in large numbers required keeping field hospital census low. Except for "lightly injured or minor illness" (especially when there is an avoidance aspect to the condition), most medical and surgical evacuations out of the theater are medically beneficial or at least not psychologically harmful.

Premature evacuation of the battle fatigued soldier (including those soldiers who initially present themselves with minor medical conditions) is a treatment approach that goes against doctrine. It creates the risk of serious social and psychiatric complications. Avoidable Post Traumatic Stress Syndrome is the most likely consequence. Thus, when viewed objectively, inappropriate evacuation may constitute medical malpractice. Returning a recovered battle fatigued soldier to his unit (and peer group) is the treatment of choice. This is true even if the soldier can not be returned until after the battle is over (or returned only during a break in fighting when the unit

is being resupplied). The battle fatigued soldier will have his own set of "war" stories to tell and he will still have the opportunity to reintegrate into the small group. The battle fatigued soldier evacuated out of the theater will never have this opportunity and is at greater risk for permanent psychological and social dysfunction.

ODS evacuation policies and limited mental health resources did not support this return to duty doctrine. Senior line and medical leaders must understand the importance of preventing unnecessary psychiatric evacuations and having sufficient mental health resources to ensure battle fatigued soldiers return to their units while they are still in the combat zone.

Conclusion

ODS represents only one part of the spectrum of conflict. It was the perfect American war because it fit our AirLand battle doctrine, our combat training, and especially our ground arsenal (Bolger, 1991). Our overwhelming success in the sky and in the desert helps ensure that the next war will be different - some other place on the spectrum of conflict with different problems requiring different solutions. Even so, the same core of mental health issues associated with soldiers, sailors, airmen, and marines engaged in combat operations will always be present. Only the range of mental health issues and the magnitude of care required will change. Now is the time to prepare for meeting these mental health requirements.

References

- Bolger, D. (1991). The ghosts of Omdurman. Parameters, XXI, No. 3, 28-39.
- FM 8-51 (1991). Combat stress control in a theater of operations. Coordinating Draft. Academy of Health Sciences, Fort Sam Houston, TX 78234.
- Knudson, G. (1991). Operation desert shield: Medical aspects of weapons of mass destruction. Military Medicine, 156, 267-271.
- Martin, J. (1991). WRAIR goes to war. USAMRDC Newsletter, July 1991. HQ, USAMRDC, Fort Detrick, MD 21702.
- Shalev, A. & Ursano, R. (1990). Group debriefings following exposure to traumatic stress. Proceedings of the Wartime Medical Services, Second International Conference (pp. 192-207). IFS Institute for Hospital Planning, Stockholm; 33 stermalsgatan; s-11426 Stockholm, Sweden.



ORGANIZATIONAL READINESS FOR TQM

J. ALAN ARENDS, Capt, USAF, MSC

Are DoD organizations ready for TQM? Or, are people distressed because they perceive they have no control over what happens to them? DoD organizations may be geared for failure of TQM because of issues and factors that have repeatedly caused new programs to fail. Organizations must prepare for change. Change must be systematic and constructed as a research design. If we simply lock-in to the rightness and righteousness of TQM without assessing "readiness" for change, we will march boldly toward a predictable future where TQM becomes a department instead of part of an organization's cultural fabric.

Facilitators of change must address organizational assessment. Ideally, assessing attitudes toward change and identifying areas needing work will occur early in structuring a new program. Weisbord (1976) developed the Six-Box Organizational Model to facilitate problem identification. Weisbord's (1976) model is a structured approach to examining organizational relationships and management. The model assesses six areas of organizational activity: purposes, relationships, leadership, structure, rewards, and coordinating mechanisms.

This study introduces a practitioner-oriented questionnaire for analyzing Weisbord's (1976) organizational relationships, adding Institutional History of Change and Current Attitudes Toward Change to indicate how changeable an organization is. Weisbord's (1976) suggested leadership factor was divided into Top Management, Middle Management, and Supervisory Team to delineate hierarchal position and provide focused assessment of leadership.

The questionnaire was administered in a field study at an Air Force hospital implementing TQM. The data about people's perceptions of the organization indicated the need for operational interventions prior to, or in conjunction with, full-scale TQM implementation.

Racial Differences in Naval Aviation Training

**Annette G. Baisden
Naval Aerospace Medical Institute**

Abstract

Throughout their histories the U. S. Navy and Marine Corps have altered their policies on minority participation reflecting societal changes in attitudes and governmental needs for military manpower. During the past three decades two major issues have shaped personnel policy and provided the impetus for a great deal of research within the military. These issues are the great force for change produced by the Civil Rights Act of 1965 and the advent of the all-volunteer armed forces concept. Most recently, leadership with the Navy has expressed increased concern about the underrepresentation of minority officers in general and minorities in naval aviation in particular. The recruiting, selection, and training processes are inextricably linked determinants of naval aviation training outcomes. This study examines historical trends in these processes, analyzes differences in success in training for minorities and majorities, and identifies factors contributing to those differences. The black population consisted of 179 students, all of the black students who entered naval aviation training during FY89-91. A matched sample of white students was selected by matching each black student with a white student on variables known to be predictive of success in training. Attrition and training data during the pre-flight and primary phases of naval aviation training were analyzed for racial differences in performances. Statistical procedures included descriptive and inferential statistics. The results showed evidence of racial differences in performance grades and types of attrition. Implications of the findings and plans for future research are discussed.

Work Attitudes Six and Twelve Months Following Implementation of a
Restrictive Smoking Policy at a Military Hospital

John A. Ballard, Ph.D.
Wittenberg University
Kenneth E. Olive, M.D.
East Tennessee State University
Guy S. Shane, Ph.D.
Air Force Institute of Technology
Major Mark E. Kain
Air Force Systems Command

Abstract

Since the 1986 report of the Surgeon General, Health Consequences of Involuntary Smoking, many organizations have implemented restrictive smoking policies. Several studies have been reported which document the impact of these policies on smoking behaviors. None of the studies reported have investigated the organizational impact of such policies. This study investigated work attitudes six and twelve months after a restrictive policy was implemented at a large military hospital. A 59 item questionnaire was used to survey over 1600 hospital employees. The initial survey had a response rate of 58%. Organizational factors examined included job satisfaction, organizational commitment, perceptions of productivity, attitudes toward smoking, and attitudes toward the policy and its enforcement. Results from the six-month study indicated non-smokers had significantly higher organizational commitment, but there were no differences in job satisfaction. Smokers, however, did perceive the policy as being enforced more stringently than did non-smokers. Over 70% of the respondents indicated a decrease in productivity as a result of the smoking policy. This finding is consistent with those corporations who have opposed restrictive smoking policies because of the economic impact of employee smoking breaks. Results from the 12-month survey are also reported.

The views expressed herein are those of the authors and do not necessarily reflect the views of the United States Air Force or the Department of Defense.

**Validity of the Structural Scales of the CPI for Predicting the Performance
of Junior Officers in the U. S. Coast Guard**

**Rex J. Blake
University of Minnesota**

**Earl H. Potter III and Richard E. Slimak
United States Coast Guard Academy**

Abstract

This poster presents a practical application of the California Psychological Inventory (CPI) to the prediction of leadership performance of junior officers in the U. S. Coast Guard. The study uses Harrison Gough's recently published structural scales or "vectors" for the CPI and contrasts the predictive utility of the structural scales with that of the original scales based on "folk concepts."

For 85 graduates of the U. S. Coast Guard Academy's class of 1986, leadership performance during the first two years following graduation was summarized in an overall rating derived from regular semi-annual effectiveness reports. CPI vectors calculated from data that had been gathered during the summer of 1982 effectively predicted leadership performance and provided some support for Gough's typological "cuboid" model of personality. While the structural scales offered a relatively compact description of personality, several of the individual folk scales, notably Dominance, performed as well or better as predictors of the performance criterion. The implications of the results both for the use of the CPI and for the utility of the kind of broad dispositional concepts that are reflected in the structural scales are discussed.

The Effects of Uniform Type and Gender on Helping Behavior

C2C Tammy Carlgren
C2C Rick Cooney
C2C Mike Pine
United States Air Force Academy

Abstract

The purpose of this study was to compare the amount of helping behavior shown towards Air Force officers, United States Air Force Academy Cadets, and Air Force enlisted personnel by residents of Colorado Springs. The scenario that was used consisted of a military person standing near a payphone. The military person would stop every tenth adult passerby and ask them if they had change for a dollar bill so that he or she could make a phone call. The independent variables that we manipulated were the type of uniform worn (officer, cadet, or enlisted) and the gender of the military person. The dependent variable we examined was whether or not helping behavior was exhibited by the subjects. We hypothesized that there would be a significant difference in the amount of helping behavior shown towards cadets when compared to officers and enlisted personnel. We also hypothesized that the female military person would receive more help than the male.

**Prediction of Absenteeism, Turnover Intention, Job Satisfaction and
Performance Using Two Motivational Models: Goal Setting and Job Design**

Abstract

Capt Terri L. Coccia

**Air Force Occupational Measurement Squadron
Randolph Air Force Base, Texas**

Researchers have studied the effects of Locke's motivational theory of goal setting on absenteeism, turnover, job satisfaction, and job performance. They have also studied the Hackman and Oldham job design theory of motivation along with variations of this model, and its relationship to absenteeism, turnover, job satisfaction and job performance. These two motivational theories have not been compared in a single study. The current study gathered data from a large DOD science and technology organization and compared the ability of these motivational models to predict absenteeism, turnover intention, job satisfaction and self-rated job performance. Overall, the job design model was a better predictor of absenteeism, job satisfaction and self-rated job performance than the goal setting model. The goal setting model, however, was slightly better than the job design model in predicting turnover intention.

Developing Learning Skills:
The Effectiveness of a Videotaped Course

C2C Richard Cuervo
United States Air Force Academy

Abstract

The purpose of the study is to see if there are any significant improvements in student performance due to the viewing of the videotape, Where There's A Will There's an 'A', developed by Professor Claude Olney of Arizona State University. Twenty students enrolled in the Behavioral Sciences introductory course in psychology will be randomly assigned to one of two groups: one group will view Olney's videotape, the other will not go through any treatment and will serve as a control group. Student performance in the class will be assessed before and after treatment. In addition to the experiment assessing the effectiveness of the videotape, pilot data will be collected on students who have completed the United States Air Force Academy's learning skills program to determine if a more complete longitudinal study of the videotape and learning skills program is justified. Factors such as training time and academic improvement will be considered.

Pressure Suit Design and Extra Vehicular Activity

C2C Crista Davis
United States Air Force Academy

Abstract

Pressure suit design with respect to extra vehicular activity has been a past, present and future point of interest in the Human Factors field. There are current studies going on evaluating two types of pressure suits: one hard and one soft, to concentrate on the pros and cons of each suit design with the astronauts interest in mind concerning survivability, mobility and maintainability in depth. The purpose of this project is to evaluate these particular tradeoffs, with specific reference to extra vehicular activity. The project will concentrate initially on the important aspects of the design of the suit, focusing on the gloves and the tasks the astronauts must perform. The project will go on to evaluate other aspects of the suit design that specifically affect the astronauts when performing extra vehicular activity. This project will also include a discussion concerning the reaction of the astronauts who have worn the pressure suits. This evaluation will be accomplished with the use of a subjective questionnaire which will be narrowed to concentrate specifically on extra vehicular activities, including the negative and positive aspects, with specific respect to survivability, mobility and maintainability. The questionnaire will be drawn up from reading previous studies on the subject as well as watching astronauts actually performing extra vehicular activity on tapes. The subjects that will be concentrated on will be those astronauts who have had training specifically geared to extra vehicular activity. Through this subjective questionnaire, coupled with personal interviews, possible new ways to approach the design of the pressure suit will give innovative insights for a redesign of the pressure suit.

Hemispheric Lateralization in Left Handed People

C2C Daniel C. Engberson
United States Air Force Academy

Abstract

Hemispheric lateralization has been shown to have significant impacts on task allocation and performance in right handed people (Polson and Friedman, 1988). Observed physical differences in the corpus collosum in left handed people (Witelson, 1985), however, present the possibility of hemispheric lateralization characteristics endemic to this population. This study follows Polson and Friedman's research, differing only in subject selection. Where they observed only right handed people, this study focuses on left as well as right handed people. The tasks are the same: a visual verbal memory loading task presented centrally to each subject. This presentation is followed by a retention interval, during which a secondary, or target task is presented. The target task is two pronounceable nonsense syllables presented laterally with the subject asked to determine whether they were the same or different. It is predicted that there will be less performance decrements in both tasks relative to their single task levels, and less or no tradeoffs between tasks than was found in right handed people due to increased and/or more efficient communication between the hemispheres.

References

- Polson, Martha C. and Alinda Friedman. "Task-Sharing within and between Hemispheres: A Multiple-Resources Approach." Human Factors, 30 (1988): 633-643.
- Witelson, S.F. (1985). The brain connection: The corpus callosum is larger in left-handers. Science, 229, 665-667.

Effects of Contextual Cue Degradation and Experience Level on Performance

C2C Dayne R. Fentress
C2C Janelle K. Viera
United States Air Force Academy

Abstract

The impact of context on learning has been well documented: subjects perform best when recall is under the same context as the learning (Tulving & Thomson, 1973, Smith, Glenberg, & Bjork, 1978). There is recent evidence that contextual cues are essential to expert performance (Beach, 1988). This study examines the effects of contextual cue degradation and experience level on performance. Subjects will be asked to rate themselves as experts, intermediates, or novices based on previous experience and proficiency with a specific video game. Each subject will play the game four times, once with no alterations (control), and then once for each cue degradation condition. We will compare the effects of degradation of color, sound, and a combination of both factors, to a baseline in which no alteration of visual or auditory cues is conducted. The order in which these conditions are presented will be counterbalanced to minimize the effects of practice and fatigue. The game scores of the subjects will be used as the performance measure. Based on previous findings, we predict that the alteration of visual and auditory cues will effect the performance of experts more than novices. The effects of cue degradation on the intermediate level, however, are more difficult to predict.

References

- Beach, K.D. (1988). The role of external mnemonic symbols in acquiring an occupation, in M.M. Gruneberg, P.E. Morris, & R.N. Sykes (Eds.), Practical aspects of memory: Current research and issues, Vol 1: Memory in everyday life. New York: Wiley.
- Smith, S.M., Glenberg, A. & Bjork, R.A. (1978). Environmental context and human memory. Memory & Cognition, 6, 342-353.
- Tulving, E. & Thomson, D.M. (1973). Encoding Specificity and retrieval in episodic memory. Psychology Review, 87, 352-373.

Analysis of Requisite Abilities of Air Traffic Control Specialists:

Toward An Effective Screening System

Gerald D. Gibb, Ph.D

Marvin L. Smith, M.S.

Neil Swindells, B.S.

Dieter U. Haney, B.S.

Embry-Riddle Aeronautical University

Abstract

The primary purpose of this work is to identify systematically the requisite abilities for success in the air traffic control specialist training environment. The results are being used to direct test development efforts in constructing an experimental selection battery. Major components of the study include findings from a literature review, training environment analyses, subject matter expert Delphi workshops, and a cognitive task analysis. The results thus far indicate that several cognitive domains appear to offer potential for success in predicting air traffic controller training performance. Primary abilities are time-sharing, working memory capacities, spatial processing, and mathematical reasoning. Current efforts are focused on developing cognitive tests for these abilities. Because of the dynamic and complex nature of the instruments needed to assess these abilities, task development is being conducted on computer-based systems.

Cigarette Smoking during Operation Desert Storm

**LTC Robert K. Gifford, Ph.D., U.S. Army
Department of Military Psychiatry
Division of Neuropsychiatry
Walter Reed Army Institute of Research
Washington, D.C. 20307-5100**

**1LT Richard T. Keller, B.S.N., U.S. Army
MAJ Ernest Degenhardt, M.S.N., U.S. Army
Department of Nursing
Fitzsimons Army Medical Center
Aurora, CO**

Soldiers receiving medical care or physical exams at the 85th Evacuation Hospital in Dhahran, Saudi Arabia, in May and June of 1991, filled out a questionnaire survey that included questions about their cigarette smoking habits, an item asking their feelings about remaining in the Army, a unit cohesion scale, a hardiness scale, and the Brief Symptom Inventory (BSI). Of the 384 respondents, approximately half reported having smoked either before or during the deployment, with the remainder reporting never having smoked. Among those who had smoked, three-quarters reported having increased their amount of smoking, with 41.9 percent characterizing the increase as "a lot more" and 33 percent as "a little more." 15.2 percent reported smoking the same amount, and only 10 percent stated that they had decreased their cigarette consumption (4.2 percent "a little less" and 5.8 percent "a lot less"). Preliminary analyses suggest a relationship between change in smoking behavior and intent to get out of the Army: those who reported that service in Southwest Asia (SWA) had made them more likely to leave the Army also increased their smoking "a lot." Soldiers who had smoked in SWA tended to expect that they would decrease the amount they smoke when they returned home, suggesting that smoking levels were seen as influenced by stress or other situation-dependent factors. Soldiers who had increased their cigarette consumption "a lot" while in SWA reported more symptoms on the BSI than did soldiers who never smoked, smoked the same amount, or smoked "a little" more or less. Soldiers who had increased their smoking more tended to be less bonded to their units' leadership, as measured by the vertical cohesion component of the cohesion scale. Hardiness was positively correlated with cohesion and negatively correlated with BSI scores. The relationship of hardiness to smoking was complex, and investigation of the relationship between these two variables continues.

Teams - A Technique of Formation

Jay W. Gould III

Abstract

Those corporations which have most successfully implemented the concepts of Total Quality have in common the strength of their teams. The formation of a team or high performance work group occurs when the team members learn to trust in each other's competence and develop strategies to compensate for the team's relative strengths and weaknesses. This paper offers a technique which may have potential to be used to reduce the cycle time of team formation.

Theory

The Learning Style Inventory developed by David Kolb presents a concept of strength, excess, and deficiency (1) for four learning styles: divergence, assimilation, convergence and execution. These learning styles are plotted on a scoring chart (2) in accordance with a rank order test (3). When each individual team member plots their test scores and compares them to the strength and weakness chart, each member can gain a better understanding of their individual abilities to accomplish learning in all phases of the Shewhart cycle (4).

A composite chart of all team members has potential to illustrate how balanced the team is in its learning behavior (5). Based on this information, the team may identify strategies for achieving balance: either through the addition of personnel of the required missing learning type(s), or through the development of compensating learning styles to overcome the deficient learning styles.

The development of a team through its stages of formation (forming, storming, conforming and performing) is learning. It is hypothesized that use of the Kolb Learning Style Inventory and identification of a team's method of deficiency compensation will help teams to progress through the stages of team formation more rapidly.

It is further hypothesized that use of the FIRO-B to screen team members can also help teams to progress more quickly through team formation. Identification of the willingness to bring others into a team and the willingness of members to be part of a team could then be used to adjust or compensate for unbalanced team composition.

External Validity in Performance Measurement Research

LCDR F. Douglas Holcombe, MSC, USN
Naval Biodynamics Laboratory

Abstract

The nature of research in a laboratory leads to an abundance of emphasis on the internal validity of the research design and on the reliability of measures. Yet the true importance of research in a military setting lies primarily with the relevance of findings in solving existing problems, its external validity. Conventional wisdom includes the notion that resultant validities represent trade-offs: stronger external validity results from "giving up" some strength in internal validity. It is contended herein, however, that through proper planning for the utility of results, external validity may be greatly improved with no loss of internal validity.

This study addresses the problem of measuring job performance in specific military tasks in such a way as to provide for maximum internal and external validity of results. An outline of a major research project currently underway at the Naval Biodynamics Laboratory is reviewed and the research design is presented as a model for strengthening the external validity and usefulness of results.

Typical tasks required of sailors aboard U.S. Navy ships are characterized in terms of a human performance taxonomy which provides a listing of basic human abilities. Each task is composed of one or more of the listed abilities. The purpose of working with the abilities level of analysis is that they are amenable to manipulation and experimentation. The purpose of working with the task level of analysis is that task performance is the key issue of the research.

Navy Recruit Psychological Screening

Imelda Idar, Ph. D.
Office of the Chief of Naval Operations
Captain James A. Scaramozzino
Bureau of Navy Medicine and Surgery

Abstract

In FY89, the Navy launched an aggressive campaign to reduce first term attrition using the principles of Total Quality Management. The goal of that campaign was to provide the fleet (the user) with a sailor who had the highest potential for supporting and maintaining fleet readiness. To that end, an analysis of attrition was conducted which revealed that the largest proportion of members separated were diagnosed as having psychological disorders. As a result, a psychological screening process was implemented at the Recruit Training Commands which identified those recruits with the highest propensity for attrition for psychological reasons. This process identifies them early in an equitable, quantifiable and consistent manner, precluding the incursion of unnecessary training costs.

The N-AFMET (Navy-Air Force Medical Evaluation Test), consisting of three phases, was implemented in October 1991. Phase I is a biographical survey which is administered to all incoming recruits. On the basis of the recruit's responses, he/she may continue in training or is referred to Phase II. That phase assesses the recruit's personality structure using the NEO Five Factor Personality Inventory and a Structured Recruit Interview. When indicated, the recruit is then referred to Phase III where a clinical mental health evaluation is conducted. Final disposition of the individual is resolved at that point.

This poster session provides descriptive data on the process, on those screened and discharged, on issues of gender bias, the value of short term counseling, impact on initial skill training attrition, and policy issues.

Schooling and Training Approaches Measurement Project (STAMP)

Abstract

Capt BJ Jones

**Armstrong Laboratory/ Human Resources Directorate
Brooks Air Force Base, Texas**

The Schooling and Training Approaches Measurement Project (STAMP), was developed in response to substantial deficiencies in the areas of curriculum design and automated instructional systems. Little or no empirical evidence is available from past instructional design research that provides principled methods for developing effective automated instruction. STAMP will be a large-scale, long-term project designed to investigate theoretical cognitive approaches in automated education and training as well as provide empirical documentation of well-specified curriculum approaches for distinct desired learning outcomes.

Armstrong Laboratory intends for the STAMP Laboratory to serve as a national testbed for instructional theory. In addition to instructional approaches proposed and developed by STAMP personnel, learning theorists of nationally recognized merit will be invited to participate in collaborative research efforts. The general investigation approach is to: (1) identify a taxonomy to guide development of a set of criterion tasks, (2) for each task, develop a set of automated instructional systems (tutors), (3) empirically evaluate each tutor, (4) investigate theoretical cognitive approaches, and (5) identify a single best instructional approach for a given knowledge type. To support the broad goals of STAMP, the computer laboratory will consist of 30 training delivery microcomputers and assets to support 50,000-plus subject hours per year.

Learning Skills Center Survey

C1C Tiffany N. Jordan
C1C Valerie J. Matthews
US Air Force Academy

Abstract

The Learning Skills Center (LSC) at the United States Air Force Academy (USAFA) serves to instruct cadets regarding methods to improve their study time, learning strategies, and organization skills. A survey was designed and administered to cadets and learning skills advisors to assess their attitudes and viewpoints of the study skills program. Analysis of the survey data showed cadets valued instruction related to time management, motivation, and student attitude. Furthermore, cadets reported overall satisfaction with the course materials and computer software products used in the program. Also, most cadets perceived the image of the Center favorably after participating in the program. One cadet had a "before" perception that the LSC is "a place for the academic speds to go" and "after" as "a place to develop your knowledge about learning -- good for all [cadets]." In summary, the survey study has detailed the current status of the USAFA Learning Skills Center. Although the current environment is positive, a few changes in cadet and faculty awareness are suggested to improve the Learning Skills Center.

History of Airpower Videodisc Program Evaluation

Jerry Lautenschlager
Patrick Patterson
Dean Orrell, Capt, USAF
United States Air Force Academy

Abstract

Evaluations based on Human-Computer Interface (HCI) guidelines are a growing part of the Human Factors field. This project is to evaluate the user interface and software compatibility of The History of Airpower Videodisc Program to be used as a multimedia teaching aide in the History Department. The project will entail an initial evaluation of a Macintosh version of the program, the results of which will be implemented in the final design of an IBM program. The main evaluation will be done on the IBM version of the program. Because there is a lack of evaluation guidelines for multimedia software, modifications will be made to existing evaluative procedures. The necessary modifications will be made to the Questionnaire of User Interface Satisfaction (QUIS) 5.0, an existing analysis guide. Once the modifications have been completed the testing will then begin on human subjects; this will involve hands-on interaction by various users, with differing user abilities. The subjects will be video-taped as they are performing the various tasks associated with the program. The preliminary video tape in addition to the subjective data achieved from the modified QUIS will be combined for an overall assessment of the programs compliance with existing HCI guidelines. The final aspect of the project will be a report to include our assessments of the user interface, software compatibility, and possible improvements to the program.

The Effect of Feedback Content on Group Practice Behavior
and Individual Achievement

Victoria Leitham
University of Southern California
Department of Educational Psychology and Technology

Michael M. Whyte
The United States Air Force Academy
Military Studies Division

Abstract

In contrast to the Iraqi fighting units during the Gulf War, American soldiers demonstrated how a sense of teamwork and mutual reliance can make the difference between success and defeat. On the educational front, teachers have traditionally discouraged teamwork, since exchanging information with others was considered cheating. Following a recent government study of industry, teachers are now encouraging students to work together. The purpose of this study is to examine how military personnel can benefit from a group-learning situation and how content of feedback affects practice behavior and achievement. Training will be delivered through a Computer Based Training program (CBT) with selected video segments that model good and poor behavior of leadership, team building, etc. There will be two phases in each treatment: a group session and an individual testing session. Fifty USAFA CW staff, faculty, officers, and cadets will be randomly selected and assigned to two treatment groups. Treatment one contains lesson, practice, and review. Feedback to practice questions will consist of explanations for correct and incorrect choices. Treatment two will be the same as treatment one except that explanations for incorrect choices will be knowledge of results only. On-line data will be collected for both group and individual sessions. This will include practice and test decision time, order of choice in the try-again procedure, module completion time, and achievement data for both group practice and individual testing. An attitude survey will be evaluated to determine how individuals benefit from the group learning situation. Information collected will provide a cost-effective alternative to individual CBT and aid in the further development of group training.

Assessing Test-Retest and Inter-Rater Reliabilities of the Worker Rehabilitation Questionnaire (WRQ).

David D. Robinson
Worker Rehabilitation Associates, Inc.
Boulder, Colorado

Kristie Prue
Colorado State University
John Howes
Colorado State University

Abstract

Hundreds of thousands of positions, representing 2,515 *Dictionary of Occupational Titles* (DOT) codes and about 93% of the jobs in the U.S. economy, have been described on the *Position Analysis Questionnaire* (PAQ). The PAQ is a standardized job analysis instrument consisting of 195 generically stated work behaviors. Data from PAQ job analyses are kept in what may be the largest and most complete job data bank in existence.

The *Worker Rehabilitation Questionnaire*, (WRQ) uses 150 job elements from the PAQ to estimate the capabilities and tolerances of individuals to perform the work behaviors, based on reports from medicine, occupational therapy, psychology and other disciplines.

The system matches estimates of individual capabilities and tolerances with characteristics of occupations in the PAQ data bank and prints a list of the closest matches. Job elements related to the individual's impairments can be coded to prevent the system from listing occupations with requirements beyond individual limits.

Forty-five vocational rehabilitation counselors employed by a state rehabilitation agency used the WRQ to rate the capabilities and tolerances of forty-five clients. The clients self-rated their capabilities and tolerances. Their spouses, parents or other concerned adults also rated these clients. Distributions of test-retest and inter-rater reliabilities were plotted and statistically summarized.

Comparisons of Combat Stress Reactions in
Two Soldier Populations During
Hostilities in the Gulf War

Major Robert R. Roland, Psy.D.

US Army Special Forces Command Psychologist

Abstract

The literature on Combat Stress Reaction (CSR) and Post-Traumatic Stress Disorder contains very limited information collected within actual field environments during wartime operations. The Persian Gulf War provided a unique opportunity to gather in-vivo data and update the existing epidemiological knowledge and etiological formulations on human stress reactions.

This presentation reviews information collected on the severity, incidence and types of stress reactions experienced by combat soldiers. All study participants were assigned to specially trained military units in Southwest Asia during Desert Shield, Desert Storm or Provide Comfort. Comparisons are made between intensively trained combatants and a similar sample of less well prepared "support" personnel who encountered similar wartime hardships. The roles and functions of proposed moderating variables, which limit the extent and severity of CSR and diminish the probability of PTSD, are outlined.

The author highlights and discusses anecdotal accounts of pre-hostility preventative strategies and on-site interventions reviewing several specific treatment cases. Follow-up research and areas for future study are proposed.

Causal Analysis of Leader Behaviors

Richard S. Tallarigo
USAF Academy

Abstract

This study investigates the causal impact of perceived leader behaviors on a variety of organizational criteria. The data are drawn from surveys of the United States Air Force Academy cadet wing. The leader behavior model under study includes five general components: leader behavioral style, organizational practices, attitudinal reactions to leadership, attitudinal reactions to the organization, and organizational performance. Leader behavioral style consists of three indicators: individualized consideration, task orientation, and leadership demeanor. Organizational practices includes four indicators: squadron goals and responsibilities, class relations, equity, and communication. Attitudinal reactions to leadership and to the organization are indicated by three items measuring satisfaction with immediate supervision and three items measuring satisfaction with the organization, respectively. Organizational performance is indicated by squadron end-of-year ranking in three areas of performance: military, athletic, and academic. The objective is to determine which causal model, among several competing models, best explains the impact of leader behaviors. This study is a follow-on to previous research (in press) suggesting that perceived leader behaviors have direct effects on attitudinal reactions to leadership, and indirect effects on attitudinal reactions to the organization. Results could have implications for policies and practices regarding cadet leadership development.

Factor Structure Comparisons of Behavioral
and Trait Assessments of Leadership

Richard S. Tallarigo
United States Air Force Academy

Abstract

The long-term objective of this research is to develop a method for empirically translating personality trait descriptions into behavioral descriptions. There are several advantages to such a translation: (a) it provides multi-method measurement of latent constructs permitting assessment of convergent validities; (b) it provides more options for individual assessment, possibly enhancing the face validity of assessment devices; and (c) because trait ratings are not as focused as behavior ratings, they may be less dependent on the observation opportunity constraints imposed by behavioral observation scales. Data were obtained from cadets at the United States Air Force Academy (USAFA) using the Adjective Check List (Consulting Psychologists Press, 1984) and behavioral items from USAFA cadet leadership assessment instruments. From results of analyses conducted so far, it appears that the factor structures of ratings from these two types of instruments contain many similarities. This factor similarity may provide the first step in developing the empirical translation of traits and behaviors. If the factors represent similar latent constructs, then the traits and behaviors may be regarded as similar indicators on those factors. This study will produce the traits and behaviors associated with each factor for comparison purposes. An outline for future research and development in this area will be suggested.

Total Quality Management and the
Training Philosophy in Cadet Squadron Eleven

C1C Sterling E. Tree
C2C Steven A. Higgins
Capt. Jay A. Horn
Lt. Col. Jeffrey S. Austin
United States Air Force Academy

Abstract

The traditional training philosophy and program at the United States Air Force Academy is rigorous, challenging, and intense. However, it could provide more growing opportunities to the highly internally-motivated cadets. Using Total Quality Management philosophies, we've altered the training philosophy to create an environment which is still very challenging, rigorous and intense, yet provides opportunities for higher levels of excellence, personal growth, and pride. Our main goal was to start using TQM principles and concepts in training to provide better instruction and experiences to the cadets.

Although TQM is a new philosophy in the management community, some of the basic concepts have been around for many years. For example, in the Contrails handbook, one of the quotes to be memorized is by General George S. Patton from his book, War As I Knew It: "Never tell people how to do things. Tell them what to do and they will surprise you with their ingenuity."

The main focus has been to give as much responsibility as possible to each individual cadet. The cadets have not let us down. We are discovering they are very capable and willing to accept more responsibility than they are given through the current regulations. The lack of a challenge is very demotivating to those searching for new obstacles to overcome.

The poster session will show the changes we've made to the current training system, and how it has affected the performance, motivation, and overall attitude of the Cadets in Eleventh Squadron.

Panel

Sexual Harassment in the Military

Robert H. Faley, College of Business Administration
Kent State University

Dan Landis, Director of the Center for Applied Research
and Evaluation, University of Mississippi

Robert E. Neibuhr and Sharon L. Oswald, College of Business
Auburn University

Mickey R. Dansby, Lt Col, USAF (ret.)

Lt Col Philip A. Irish, III, DEOMI Director of Research

This panel will present analyses of two separate surveys each of which relate to the frequency of sexual harassment surveys in the military services. The first database is the 1988 DoD Survey of Sex Roles in the Active-Duty Military which mandated by then Secretary of Defense Frank Carlucci. This database includes the responses of approximately 20,000 active-duty personnel across all four services.

The second database results from the administration the Military Equal Opportunity Climate Survey (MEOCS) which, to date, includes the responses of over 15,000 active-duty personnel, primarily in the Navy and Army. This survey includes many items relating to sexual harassment.

During this panel, the participants will summarize the relevant parts of the respective surveys. They will then focus on the similarities as well as differences in the results. Finally, the implications for training will be specifically addressed.

Panel Presentation

Preliminary, Partial Estimates of the Annual Dollar-Value of Overall Lost Productivity Due to Sexual Harassment in the Active-Duty Military

Robert H. Faley
College of Business Administration
Kent State University

Preliminary, partial estimates, using well accepted economic assumptions, of the productivity-related costs of sexual harassment in the active-duty military are derived. These estimates indicate that in 1991 dollars, the annual cost of sexual harassment is over \$40,000,000 with over 90% the result of impact on enlisted personnel. The implications of these and other results for the way the military will operate in the future are discussed.

Panel Presentation

MEOCS and Sexual Harassment

Dan Landis, Director of the Center for Applied Research
and Evaluation

Lt Col Philip A. Irish, III, DEOMI Director of Research
Mickey R. Dansby, Lt Col, USAF (ret.)

The Military Equal Opportunity Climate Survey (MEOCS) contains 10 items which describe sexual harassment/sex discrimination behaviors. In this presentation, we relate those items to measures of job satisfaction, commitment, and perceived work group effectiveness as well to a number of demographic variables (e.g., ethnic group, paygrade, and rank). Results indicate that the occurrence of these behaviors is particularly deleterious to enlisted black women in the lower ranks. These results are discussed in terms of their implications for equal opportunity training in the military services.

Panel Presentation

The Influence of Workgroup Composition on Sexual
Harassment Among Military Personnel

Robert E. Niebuhr
Sharon L. Oswald
Department of Management
Auburn University

Data from the 1988 DoD Survey of Sex Roles in the Active-Duty Military indicated that patterns of sexual harassment differed across two aspects of workgroup composition: females in traditional male jobs and females in male-dominated work units. Results also indicated that other individual and situational variables provided significant interaction with work-group composition in explaining sexual harassment incidents.

The Role of Total Quality Management
in Getting the Anti-Personnel Obstacle Breaching System
to Operation Desert Storm

Felipe A. Garcia
Naval Coastal Systems Center
LTC Stanley H. Holgate, Ph.D.
Carolyn K. Bense, Ph.D.
Natick Research, Development and Engineering Center
Steve Jones
Naval Ordnance Station, Indian Head

After the invasion of Kuwait in August 1990, the Iraqi army emplaced millions of anti-tank and anti-personnel mines to slow the advance of the multi-national forces arrayed against them. The Anti-Personnel Obstacle Breaching System (APOBS), one of several countermine development efforts under the sponsorship of the U.S. Marine Corps, was being developed by the U.S. Navy with testing support from the U.S. Army. It was in advanced development testing with an initial operational capability slated for September 1993. The engineering development and testing efforts, under the leadership of the project officer at the Naval Coastal Systems Center, were being managed using total quality concepts in order to continually improve the APOBS for the customers, Marines and soldiers in combat. Much of the coordination among participating agencies was accomplished through electronic mail. Because team building had taken place early in the acquisition cycle, the APOBS team was able to compress the schedule for critical development and user interface testing sufficiently to field the APOBS for limited use in Operation Desert Storm.



The Total Quality Management Model

Colorado

Department of Personnel

State of Colorado

Sue Huang, Ph.D.

92-17692



Abstract

A panel of three members will present the Total Quality Management model recently designed for the Department of Personnel, State of Colorado. This model was selected to increase work quality and productivity of the Department and to exemplify Governor Romer's commitment to quality work within state government.

Under the direction of the Executive Director, Department of Personnel, a group of eight employees was invited to serve as the Quality Council to install a total quality management (TQM) program for the department. With the guidance of an outside consultant, the council designed a TQM model for the Department. A brainstorm session was held to generate ideas to improve the operations within the Department. A description of the TQM model was presented to all employees within the Department. All necessary training, e.g., teamwork, time management, meeting management, will be provided to employees when necessary. It is hoped that after the program is launched, customer services, internally and externally, will be improved.

Within the Department of Personnel, Total Quality Management is defined as a philosophy and set of techniques for the organization to work together to better meet its customers' needs and to be more productive and cost effective. This general concept has been introduced in the state government at various levels. Many state agencies have initiated efforts toward this mission since 1989.

The major TQM components emphasized by many state agencies are (1) teamwork with heavy emphasis on working together: training in teamwork, interpersonal dynamics and meeting management has been offered; (2) customer focus: several state agencies have done customer surveys to determine customer needs and to design tasks to improve customer satisfaction; (3) work systems analysis: state agencies analyze their work systems and develop a database for decision making and management by fact; and finally (4) continuous improvement as well as long range strategic thinking. The Department of Personnel has adopted these components for its effort and believes it will improve the department's quality services continuously.

Underlying this approach is the belief that the whole is greater than the sum of its parts. The overall outcome reached by all employees in a collaborative manner will be far greater than the employees individually could make. Consequently, it is made clear to all employees that they are players in this TQM model. Each employee may be involved at one or more levels in this activity; these are (1) quality council membership, (2) quality task teams membership, and (3) individual employee contribution.

The roles of the quality council are (1) to steer the program, (2) to identify departmental tasks and (3) to set up departmental level task teams. The roles of the quality task teams are (1) to work on departmental level or division level tasks, (2) to review areas for improvement and (3) to recommend ways to improve service quality. Last but not least is the employee level participation in which the employees serve as either council members or task team members and constantly improve ways to do their own work.

Specifically, the quality council is made up of eight employees representing all divisions and all levels of job classifications. All eight employees will have equal authority on issues and actions taken at the quality council level. Half of the members will serve a 6-month term, while the other half will serve a one-year term in order to provide continuity for the council. The Executive Director of the Department will serve as a permanent member. All council members will work on a consensus basis to reach recommendations for all issues discussed. The council will identify possible tasks or projects based on the inputs received. The priority of tasks or projects will be determined. Potential task team members will be identified and invited to participate. General expectations will be outlined to the selected task team members. After that, the task team members will work independently.

A task team only works on a task or a project for a certain period of time. Together, task team members analyze the task or the project and work on solutions for the selected task. After completing an assignment, the task team will communicate its recommendations to the quality council, or to the department management team or to all employees within the department, if necessary. Finally, the team will have the authority to see that the recommendations are implemented. Any follow-up, if necessary, will be relayed back to the quality council for further improvement.

If employees are not serving on a task team or the quality council, they still are able to make their contribution to this TQM effort by improving their own job performance, recording suggestions or identify problems to their managers or the quality council. Ultimately, employees are encouraged to participate in the council or any task team.

This interactive TQM model will require constant communications among the quality council, task teams and individual employees. The resources for tasks or projects in terms of people, time and funding as well as training needs will be identified with inputs from all involved parties. Data such as tasks/projects identified, ways to improve tasks, employees who participate and the outcome of the recommendations will all be kept by the quality council. Different mechanisms will be designed to assess the workability of the TQM model in the area of customer satisfaction, cost-saving and efficiency of work flow.

Conclusion

With this TQM model all employees within the Department of Personnel will be actively involved in the total quality management process. It is hoped that the goals of this program, i.e., improving customer services, producing better quality products and saving operations cost, will be achieved and continuous improvements realized.

A Comprehensive System for Minimizing
Organizational Delinquency

Robert T. Hogan, Ph.D.
Joyce Hogan, Ph.D.
University of Tulsa
Ralph M. Carney, Ph.D.
Personnel Security and Education Center
Janice S. Houston, M.S.
Personnel Decisions Research Institutes, Inc.
Chase P. McCown, B.S.
Wright Laboratories
Frederick W. Gibson, Ph.D.
Gordon J. Curphy, Ph.D.
United States Air Force Academy

Abstract

Although a recent APA task force report (Goldberg, Grenier, Guion, Sechrest, & Wing, 1991) and meta-analytic results (Ones, Viswesvaran, & Schmidt, 1990) generally support the use of integrity tests for identifying counterproductive workers in blue collar jobs, it is unclear whether these findings generalize to individuals in high security jobs, such as nuclear power plant operators or individuals with access to sensitive government materials. This uncertainty is primarily due to the lack of published research regarding the efficacy of integrity testing for screening individuals applying for high security jobs. The members of this panel describe a system which combines selection techniques with an on-the-job monitoring program in order to minimize the counterproductive work behaviors of those individuals in high security jobs. More specifically, the members of this panel summarize the development and research findings concerning background investigations, integrity tests, and monitoring systems and describe the utility of using a comprehensive screening and monitoring system for minimizing organizational delinquency among workers in positions with high security requirements.

Predicting Organizational Delinquency Among
Pre-Screened Subjects: A Concurrent Validity Study

Gordon J. Curphy, Ph.D.
Frederick W. Gibson, Ph.D.
Kent A. Landreth

Department of Behavioral Sciences and Leadership
United States Air Force Academy

Chase P. McCown
Wright Laboratory
Wright-Patterson Air Force Base

Abstract

This study examined the potential of using integrity tests to predict organizational delinquency among subjects who have already undergone extensive background investigations. Although the criterion measures used were more comprehensive than reported in most studies of organizational delinquency, few significant correlations were found. The implications of these results and the need for monitoring systems to minimize organizational delinquency in high security jobs is discussed.

Although delinquency has been a topic that has generated much research over the years, organizational delinquency is a relatively new topic in psychology. Organizational delinquency consists of a number of counterproductive employee behaviors which have negative consequences for an organization (Hogan & Hogan, 1989). Whereas organizational delinquency consists of a broad range of counterproductive behaviors, previous research has focused primarily on detecting employee theft or substance abuse in blue collar jobs. Integrity tests have been the primary tool used for detecting these two counterproductive behaviors among workers. Furthermore, very little published research is available to evaluate whether these tests are valid predictors of organizational delinquency for high security jobs, such as nuclear power plant operators or intelligence personnel. This lack of research is unfortunate, as delinquent behavior in these positions can result in more severe organizational consequences. In addition, because background investigations are a part of the selection process for many of these high security positions, the base rate of counterproductive behaviors among these workers may be substantially lower than those in blue collar positions. Given this lower base rate, the ability of integrity tests to identify applicants likely to engage in organizationally delinquent behavior is uncertain.

Biodata, background investigations, and integrity tests have all been used to determine organizational delinquent behavior among applicants. For example, McDaniel (1988) used a sample of 10,188 enlisted personnel to investigate the utility of pre-employment drug use information for predicting unsuitability discharges from the military. Overall, indices of pre-employment

drug use correlated less than $r=.09$ with being discharged from the military. Carney (1991) conducted a study which examined the reasons why applicants were denied security clearances as a result of background investigations. Applicants who were denied clearances often had (a) poor credit histories, (b) previously used illicit drugs, (c) foreign relatives, or (d) provided false self-report information. In addition, Timm (1991) stated that the number of rejected applicants may underestimate the power of background investigations to minimize organizational delinquency, as many potential applicants with questionable histories may self-eliminate before undergoing a background investigation.

Wiskoff, Parker, Zimmerman, and Sherman (1989) looked at the feasibility of using both biographical and personality information for selecting Marines for Marine Security Guards (MSGs). They reported that certain biodata and personality scale scores were useful in predicting behavior reliability among MSGs. Although the overall correlations were fairly low ($r=.11-.35$), it is important to remember that the range was most likely restricted for the reasons associated with background investigations mentioned previously. The authors also advocated using an on-station monitoring system to further minimize counterproductive behaviors among MSGs.

In addition to biodata and background investigations, integrity tests have also been used to predict organizational delinquency among applicants. Integrity tests can be grouped into two main categories, overt and personality-oriented. Overt tests are narrower in focus, and are primarily designed to detect employee theft proneness, honesty, and/or substance abuse problems. Personality-oriented based integrity tests are broader in scope than overt tests, and assess a wider variety of counterproductive behaviors, such as malingering, tardiness, insubordination, theft proneness, substance abuse problems, and honesty. Using meta-analytic techniques, Ones, Viswesvaran, and Schmidt (1990) found that personality oriented measures were better predictors of a broader range of counterproductive behaviors than the overt tests. Based on the Ones et al. (1990) findings, two personality oriented integrity tests were used in this study. The specific hypothesis tested in this study was that the Reliability scale of the Hogan Personality Inventory (HPI) and various scales of the Profile personality inventory would be correlated with various measures of organizational delinquency.

Method

The United States Air Force Academy (USAFA) is a four year undergraduate military institution. The mission of USAFA is to prepare cadets to be commissioned officers and leaders in the US Air Force. As part of the admissions process all cadets underwent an extensive screening and review process, which included interviews, medical tests, physical fitness testing, and an extensive review of high school academic records, extra curricular activities, and any prior involvement with the police. To maintain standards and instill a sense of responsibility at USAFA, cadets are punished for poor performance or unacceptable

behavior. USAFA regulations formally specify which behaviors are punishable, and provide recommendations for disciplinary action. Examples of punishable behaviors described in these regulations include being late to or missing class, wearing uniforms out of standards, using poor judgment, driving under the influence, abusing superior-subordinate relationships, lying, or cheating. All instances of formal punishment are documented in a cadet's personal records. The punishment data was then organized into eight mutually exclusive categories. The categories included: (a) being late to or missing class, (b) alcohol related incidents, (c) abusing superior-subordinate relationships, (d) financial mismanagement, (e) negligence of duty, (f) substandard personal appearance, (g) vehicular infractions, and (h) being over the fence or absent without leave.

The individual difference measures used to predict the punishment categories listed above were the Hogan Personality Inventory (HPI) (Hogan, 1986), and the Profile personality inventory (Jones, 1988). The HPI is a 206 item, true false, self-report inventory which is designed to assess an individual's standing on seven primary, six occupational performance, and two validity scales. Of these scales, only the Reliability occupational performance scale was used in this study. Individuals with low Reliability scale scores are thrill-seeking, impulsive, hostile to authority, and insensitive. The Profile is a 186 item, Likert scaled inventory which assesses an individuals standing on 11 substantive and two validity scales. The 11 substantive scales include, Argumentative, Interpersonal Insensitivity, No Common Sense, Unstable Relations, Attention Seeking, Arrogance, Fear of Failure, Dependency, Perfectionism, Passive-Aggression, and Untrustworthiness. High scores generally indicate that an individual is hard to work with.

The sample consisted of two groups of 108 and 109 subjects each, and their demographics closely mirrored the cadet population in general. Formal punishment data was collected from the subjects' personal records for January, February, and March 1991. This data was then compared to test and questionnaire results administered in April 1991. Participation was voluntary, and anonymity was guaranteed.

Results

The man Profile scale scores were similar to those of a normative sample, however, cadets score about a standard deviation lower on the Reliability scale than a normative sample. These results indicate that the cadet sample was more likely to engage in behavior that would be considered reckless, non-conforming, or impulsive.

The primary hypothesis tested in this study looked at the correlations between the Reliability scale of the HPI and various scales of the Profile, with the eight measures of organizational delinquency. Only 6 out of the 96 correlations were significant at the .01 level, and those correlations which were significant do not have an absolute value greater than $r=.24$.

**Intercorrelations Between the Individual Difference Variables and
the Organisational Delinquency Categories**

Predictor Variables	Criterion Variables							
	Late	Duty	Veh	Sup- Sub	Fin	App	OTF	Alc
Reliability	-.04	-.04	.23**	-.06	-.04	.07	.01	-.06
Interper Insen	-.14	-.13	-.15	-.11	-.10	.08	.00	-
Argumentative	-.20	-.03	-.07	-.19*	-.03	-.04	-.04	-.06
Unstable Rels.	-.21*	-.04	-.18*	-.12	-.11	.01	-.04	-
No Common Sense	-.10	-.03	-.01	-.01	-.09	-.02	-.10	-
Attention Seeking	.00	.08	-.11	.15	.18*	-.04	-.03	-
Arrogance	.00	.10	-.11	-.07	-.04	.01	.01	-
Fear of Failure	-.01	-.02	-.08	-.16	-.05	-.01	-.06	-
Dependency	.01	-.03	-.05	-.03	-.08	-.02	.01	-
Passive Agg.	-.04	-.08	-.11	-.13	-.15	-.03	.00	-
Perfectionism	.05	.05	-.09	.02	.08	.05	.01	-
Untrust.	-.14	-.09	-.02	-.13	-.18*	.00	.03	-

*p<.01 **p<.001

Discussion

We were surprised that the Reliability scale of the HPI and the Profile scale scores were not significantly correlated with any of the organizational delinquency categories. In addition, those correlations that were significant were in the opposite direction. We believe the criteria used in this study were among the best available as operational definitions for the organizational delinquency construct, especially for pre-screened employees. However, even these indices have shortcomings. Probably the biggest problem with the punishment criteria is that they were only measures of who got caught performing some type of counterproductive behavior. They were not accurate measures of the frequency in which counterproductive behaviors were actually performed. Most likely only a small amount of counterproductive behaviors were detected, and even a smaller amount were formally punished. Although the measures in this study may have assessed only a fraction of the counterproductive behaviors actually performed, it is also possible that the actual base rate of these behaviors for cadets is quite low. Individuals likely to engage in these counterproductive behaviors may have never been accepted to the Academy, or may have self-eliminated as a result of having to undergo a background investigation. Furthermore, all cadets in the sample had been at USAFA for at least six months, and socialization effects may have made cadets less likely to engage in acts that they had already been or seen others punished for.

Moreover, six out of the eight punishment categories had less than seven documented cases. As a result, it may be that range restriction, rather than a lack of true relationship, may

be the primary reason why many of the predictor-criterion correlations failed to reach significance.

Although integrity tests scores appear to be unrelated to organizational delinquency among cadets, we do recommend their use to help minimize organizational delinquency among workers in high security jobs. Polygraph tests have been frequently used up to this point, however, their shortcomings have been well documented. It is very likely that thorough background searches may be the best predictors of organizational delinquency. Although integrity tests may only account for a small amount of criterion variance over and above these searches, given the consequences of delinquent behavior in high security jobs, any increase in the amount of criterion variance accounted for may be helpful. Furthermore, using high cutoff scores to minimize false negatives may outweigh the consequences associated with a high false positive rate with these tests. Probably the best approach to minimizing organizational delinquency would be to use a combination of background investigations and integrity tests as selection criteria, followed by an on-the-job monitoring process.

References

- Carney, R.M. (1991, August). Developing criteria-driven personnel security investigations. In R.P. Denk chair, **Personnel Security in the U.S. Department of Defense- The Criterion Issue**. Symposium conducted at the 99th meeting of American Psychological Association, San Francisco.
- Hogan, R. (1986). **Hogan personality manual**. Minneapolis, MN: National Computer Systems.
- Hogan, J. & Hogan, R. (1989). How to measure employee reliability. **Journal of Applied Psychology**, 74, 273-279.
- Jones, W.H. (1988). **Manual for the Profile personality inventory**. Unpublished manuscript.
- McDaniel, M. A. (1988). Does pre-employment drug use predict on-the-job suitability? **Personnel Psychology**, 40, 717-730.
- Ones, D.S., Viswesvaran, C., & Schmidt, F.L. (1990). Integrity testing: **Empirical confirmations and refutations using meta-analysis**. Unpublished manuscript. University of Iowa, College of Business Administration, Ames.
- Timm, H.M. (1991). Defining financial responsibility. In R. P. Denk (Chair), **Personnel Security in the U.S. Department of Defense- The Criterion Issue**. Symposium conducted at the 99th meeting of American Psychological Association, San Francisco.
- Wiskoff, M. F., Parker, J. P., Zimmerman, R. A., & Sherman, F. (1989). **Predicting school and job performance of Marine Security Guards**. (PERS-TR-90-003). Monterey, CA: Defense Personnel Security Research and Education Center.

The views expressed in this paper are those of the authors; they do not necessarily represent those of the United States Air Force, nor the United States Air Force Academy.

The authors would like to thank the Frank J. Seiler Laboratory at the United States Air Force Academy for funding this research.

Selection in Leadership: Some New Research Directions

Robert T. Hogan, Ph.D.

Joyce Hogan, Ph.D.

University of Tulsa

Dianne L. Nilsen

The Center for Creative Leadership

Richard L. Hughes, Ph.D.

Frederick W. Gibson, Ph.D.

Gordon J. Curphy, Ph.D.

United States Air Force Academy

Abstract

Three recent findings in the fields of personality and industrial/organizational psychology may have important implications for the ways leaders are selected in the future. One finding concerns the relationship of self versus others' ratings of personality with leadership effectiveness; others' ratings appear to have higher correlations with leadership effectiveness. Furthermore, methodological artifacts do not seem to play a major role in differences in the correlations generated using the self and others' ratings. Another finding concerns the relationship between cognitive abilities and leader performance under stress. Some cognitive abilities have higher validity coefficients with leader performance under high stress conditions, whereas other cognitive abilities have higher validity coefficients with leader performance under low stress conditions. The third finding concerns the use of low fidelity simulations in leader selection. Low fidelity simulations ask respondents to read various leadership scenarios and indicate what actions they would take in response to each of the scenarios. The panel members review the research related to self versus others' ratings of personality, cognitive abilities and stress, and low fidelity simulations and discuss their implications for leader selection.

Panel

Equal Opportunity Climate and its Measurement

Dan Landis, Director of the Center for Applied Research
and Evaluation, University of Mississippi
Mickey R. Dansby, Lt Col, USAF (ret.)
Paul Rosenfield, Navy Personnel and Development Center
Lt Col Philip A. Irish, III, DEOMI Director of Research
Col Ronald Joe, DEOMI Commandant

The Department of Defense has mandated that units assess the state of their equal opportunity climate. For the past two years, the Defense Equal Opportunity Management Institute has made available the Military Equal Opportunity Climate Survey as one tool to meet this requirement. In addition, the Navy, through the Women and Multicultural Research Office has developed a complimentary instrument. This panel will address the following issues:

1. What are the findings from MEOCS after two years?
2. What are the results of the Navy's Equal Opportunity/Sexual Harassment Survey (NEOSH)?
3. What is the relationship between equal opportunity climate and unit mission as well as minority/majority proportions?

Implications of these findings for training and selection will be discussed.

Panel Presentation

MEOCS after two years: Major and some minor findings

Dan Landis, Director of the Center for Applied Research
and Evaluation, University of Mississippi

Mickey R. Dansby, Lt Col, USAF (ret.)

Lt Col Philip A. Irish, III DEOMI Director of Research

The Military Equal Opportunity Climate Survey, designed as a tool for unit commanders was released for operational use in the summer of 1990. To date 95 units, primarily from the Navy and Army, have participated. These units have ranged in size from carrier squadrons (around 20 personnel) to Army Division (a sample of close to 1000). The total number of personnel who have participated is, to date, over 15,000. Results have continued to indicate good to excellent reliability and construct validity. This paper discusses these findings, points out consistent differences between minority and majority members, men and women, and assesses the status of equal opportunity climate in the active-duty services today. In addition, the implications for training of personnel is addressed.

Panel Presentation

The Impact of Unit Mission and Majority/Minority Ratio on Equal Opportunity Climate

Lt Col Philip A. Irish, III, DEOMI Director of Research
Dan Landis, Director of the Center for Applied Research
and Evaluation, University of Mississippi

Prevailing beliefs indicate that combat units should have higher morale and, consequently, higher levels of positive equal opportunity climate. As well, we would expect units with high ratio of minority to majority members to have a lower measured EO climate. Both these hypotheses were tested using the Military Equal Opportunity Climate Survey (MEOCS) database. Units were categorized in terms of their primary mission as combat, combat support, or service support. The categorization was made by personnel from the various services using consensually agreed criteria. In terms of majority/minority ratio, units were categorized in terms of their deviation from the overall DoD ratio (approximately 70 white). Units were categorized as high majority if their percentage of whites exceeded 84% (120 percent of the DoD figure) and low majority, if they fell below 56% (80 percent of the DoD figure). The hypothesis with regard to majority/minority ratio was confirmed. However, the suggestion with regard to mission type not only failed to be confirmed, but the results indicated an opposite effect. That is, combat units had the lowest EO climate and service support units the highest. The implications of these results for further studies and training doctrine are discussed.

Panel Presentation

Assessment of Equal Opportunity Climate in the Navy

Paul Rosenfield
Navy Personnel Research and Development Center
Women and Multicultural Research Office
Code 01E
San Diego, CA 92152-6800

As a result of recommendations from high-level Navy study groups, the Navy Equal Opportunity/Sexual Harassment Survey (NEOSH) was developed, administered, and analyzed by researchers at the Navy Personnel Research and Development Center in San Diego, CA during FY89-90. The goal of the NEOSH was to provide Navy policymakers with an accurate baseline measure of EO climate and sexual harassment among active-duty Navy personnel. Since the survey will be administered biennially, the EO climate portion of the NEOSH was developed from previous EO climate assessment efforts, along with original questions designed to investigate additional topics of concern to the Navy. A modular overall design was employed with items related to specific EO topics (e.g., assignments, discipline) grouped together.

The NEOSH was administered to a random sample of active-duty Navy officers and enlisted personnel stratified on racial/ethnic group and gender. A total of 5,558 completed NEOSH survey were received (response rate=60%) and analyzed.

The major results of the 1989 administration of the NEOSH were:

1. Navy personnel, as a whole, have positive perceptions of EO climate.
2. White male officers consistently report the most positive perceptions of Navy EO climate. The differences in EO perceptions between male and female officers are typically larger than between male and female enlisted personnel.
3. Blacks, particularly black enlisted females, are the least positive about EO.
4. Perceptions of fairness in discipline are clearly lower among blacks.
5. Blacks and women are more likely to feel they have to work harder to get promoted/advanced.
6. Hispanics EO perceptions consistently fall between whites and blacks and typically are closer to whites.
7. As paygrade and rank increase so do perceptions of EO climate. However, differences between enlisted paygrade levels are larger than between officer ranks.
8. While for whites the increase in EO climate with increasing rank and paygrade systematic, for women and minorities there was not a clear increase between the junior and midcareer levels. The largest increase for women and minorities was generally between midcareer and senior levels.

Medical Reserve Personnel Demobilized After Operation Desert Shield/Storm

A. David Mangelsdorff, Ph.D., M.P.H.
U.S. Army Health Care Studies and
Clinical Investigation Activity
Health Services Command (ATTN:HSN-T)
Fort Sam Houston, Texas 78234-6060

During Operation Desert Shield/Storm, more than 40,000 medical reserve personnel were activated. They were deployed nearly 50/50 between Southwest Asia and Europe and staffing military hospitals in the United States. The greatest problem in the reserve activation was the call-up was made by units. Unfortunately, entire units were not always needed. This was the first reserve activation since the Korean War.

A number of studies of medical reserve personnel were conducted. This panel will examine the results of some of these efforts in different service groups. Discussion will focus on the similarities and differences of the findings.

The participants in this panel include:

Shelley Perry and James Griffith, WESTAT, Inc.

Ludwig Uhlmann, Federal Republic of Germany

A. David Mangelsdorff, U.S. Army Health Services Command

D. Stephen Nice, Naval Health Research Center

Special USAR Populations of Concern During Mobilization

Shelley Perry, Ph.D.
James Griffith, Ph.D.
WESTAT, Inc.
Rockville, Maryland 20850-3129

The availability and reliability of the citizen-soldier are the underpinnings that determine the role of the Reserve forces in the national military strategy. While the Desert Storm deployment may have raised more questions than were answered on reliability with regard to combat arms units as opposed to combat service support, we are able to extract meaningful insights on reserve soldier availability and difficulties encountered because of sustained deployments.

There are societal and political conditions which are likely to not be replicated during subsequent deployments, but many of the personal conflicts encountered because of the separate private citizen and soldier roles will be similar. An additional reason to carefully examine the results of this experience is that most all previous research had dealt with the conflicts that occurred during regular training. Although the balance between these two roles (Citizen-Soldier) is more easy to maintain during weekend drill and annual training, ultimately the value of the reserve concept must be able to prove itself during short-notice sustained deployments.

During Desert Storm, Army Reservists were called upon to fulfill the soldier role for which they have been prepared and financially compensated during peace. They were now called upon to put their role as the soldier dominate to that of family and civilian employment obligations. Certain groups, because they were thought to be likely to encounter the most complicated and costly aspects of family and employment conditions, were a source of particular concern to USAR leadership with regard to deployment difficulties. Among these groups are:

- * Single parents, including custodial and noncustodial parents;
- * Dual-status Reservists; Reserve members married to other members of the U.S. military; and
- * Medical personnel, including registered nurses, licensed practical nurses, and physicians.

The numbers of single parents and dual-status Reservists have vastly increased since the last mobilization of Reservists. The demands of their family status have the potential to affect the entire deployment process. Because of complications over child-care arrangements, the concern is that these groups are not as prepared for deployment and may not deploy at the same rates as the rest of the unit. Unit leadership may also face an additional burden to ensure deployment of these individuals (i.e., ensuring that child care plans are valid). The ultimate effects on retention after deployment among these groups is also a concern, after they have experienced the demands of a deployment.

Just as single parents and dual-status parents were thought to be the most likely to encounter significant family-related difficulties, medical personnel in the Reserve were thought to be the group which would encounter the most severe civilian job conflicts. They were thought to be the most likely to either not deploy, or to leave the reserve upon return from deployment. In general, this expectation was the result of the perception of lost clients (patients) and financial obligations. While many of the circumstances would be typical of any small business, medical personnel, and in particular doctors, were thought to be the group greatest at risk. The concern was also highlighted because of the large number of medical personnel required in the Reserve as opposed to any other profession. Their critical role in the USAR and the One Army Concept as a whole, places particular importance on their mobilization experience and their intentions following that mobilization.

Each of these special populations received a great deal of media focus during Operation Desert Shield/Storm (ODS). Did the media overpublicize one or two vocal cases of hardship during mobilization or did the media uncover a new "social problem" for the USAR? The specific questions addressed in this paper are:

- * Did groups deploy at similar rates to the USAR as a whole?
- * Was the experience of deployment unusually stressing to these individuals and their families?
- * Do members of these groups intend to stay in the USAR after the ODS experience?

This paper will present results from an annual survey of Army Reservists conducted for the Chief, Army Reserve. An oversample of these three special populations within the USAR were included in the 1991 survey, administered during July through October. Survey returns were received from 1,648 single parents, from 6,110 dual status Reservists, and from 1,482 medical personnel, including 292 physicians, 800 registered nurses, and 376 LPNs, and among all groups, we have returns from both deployed and nondeployed groups.

Attitudes and perceptions about reserve service are examined, both before and following Operations Desert Storm/Shield. Topics of analysis will include: attitudes about the ODS experience, current career intentions and the rationale behind those intentions. Survey results will be presented in light of implications for future recruitment and retention of medical professionals in the USAR.

Operation Desert Shield/Storm:
Army Reservists in need of financial support?

Ludwig Uhlmann, Diplom-Psychologe
Federal Republic of Germany

AD-P006 961


Based on a survey conducted on Army reserve personnel who were deployed stateside or overseas during Operation Desert Shield/Storm (ODS/S), the purpose of this study is to obtain guidelines for how to improve social services for reservist soldiers and their dependents or significant others in times of crisis or war.

Questions to be answered were: Are there significant differences between certain groups of respondents as depicted by their life-style patterns and how they perceived and coped with stressors due to ODS/S. If so, what would be recommended for optimizing support to individuals who really need it.

Stressors due to crisis, disasters, or war are seemingly unlimited; financial resources to smooth stress peaks are definitely limited. So there is a strong need to focus on those individuals who are really suffering from the impact of stressors and need intervention and others who could more or less be left on their own after a first encounter with social workers or mental health care providers. Typically reservists are exposed to their civilian dominated social demands as well as to the pressure induced by duties in their military environment when activated. So the acquisition, evaluation, and interpretation of the data must reflect these two dimensions which may lead to a different approach to mental health care intervention as in comparison to active military professionals. The question is which information about the individuals could be regarded as most helpful to separate critical cases from others. To obtain relevant information for this purpose involves quantitative and qualitative analysis of the survey data.

Method

Subjects The sample consisted of 213 Army Reserve personnel who were called to active duty in support of the Persian Gulf Conflict. The soldiers belonged to two units: 105 to the 5501st USAH, St. Paul, Minnesota, and 108 to the 145th MEDSOM, Texarkana, Texas. All personnel served either in Saudi Arabia or the United States. When they were inactivated, both groups outprocessed at Fort Sam Houston, Texas, where the survey was administered in group sessions in April and May 1991. The questionnaire was purely voluntary and strictly anonymous.

Procedure The survey consisted of 62 close-ended questions; scales were dichotomous or Likert type scales. The questions were related to 244 variables. These variables were grouped into a descriptive, categorical portion, a portion representing stress-perception and stress-symptoms, and a portion reflecting stress-coping strategies or skills. The first group served as "independent", the remaining two as dependent variables in the statistical designs.

The first group of variables was used for a set of descriptors or properties of respondents which could reveal - according to their qualities - different relations between them and the dependent variables and hence contribute to classifying individuals with respect to their stress tolerance or proneness.

92-17694



These variables and categories were:

Age groups	<-25/ 26-35/ 36-45/ >45
Gender	F/ M
Marital Status	single/ married
Pay Level	small/ medium/ top
Education	lower/ higher
Years of Service	<-4/ 5-12/ >13
Months of Service ODS/S	<-3/ 3->6/ ->9/ >9
Location	US/ Saudi/ overseas
Finances	Loss/ NoChange/ Gain
Employment Spouse	Yes/ No
Children	Yes/ No
Number of Children	1/ 2/ 3->
Num of Chldrn in School	1/ 2/ 3->
Single Parent	Yes/ No

Quantitative analysis of data included nonparametrics and ANOVA or MANOVA procedures when appropriate. Qualitative analysis of data was performed on the basis of descriptive statistics and profile comparisons for extreme groups.

Results

The 213 respondents ranged in age from 19 to 58 and averaged 31.8 years in age. Thirty-two percent (N=69) were female and 68% (N=144) were male. Roughly half (55%) were married or remarried and a considerable portion (42%) hold college or advanced college degrees. The sample consisted of 155 (73%) enlisted personnel and 58 (27%) officers. These soldiers had been in the military from 1 to 28 years with an average of 9.17 years. Duration on active duty due to the Persian Gulf Conflict ranged from 2 to 62 weeks, average is 19 weeks. Half (53%) performed their duty in Saudi Arabia, while the other half remained in the United States. Forty-one percent suffered financial losses, no change in their financial situation was incurred by 19%, and 40% stated financial gain.

Overall, the self rated health level with respect to somatic manifestations after active duty was not regarded as problematic by the majority (90%) of the respondents. Differences between respondents with respect to all of the grouping variable properties (see above) and the self rated health level were not significant. All respondents except two - regardless of their experienced health problems - reported good/very good military job performance during their deployment. Similarly the extent of perceived stress level (low vs. high) did not show significant influence on job performance ratings, neither did the high scorers on symptoms listing vs. low scorers show significant differences.

The range of the soldier's extent of perceived stressors was 4.40, minimum 2.20, maximum 6.6, mean 4.31, median and mode 4.00, variance 0.63, and standard deviation 0.79.

One-way ANOVAs performed on the soldier's extent of perceived stressors testing for differences between the factor levels of the descriptor variables produced significant main effects for age groups, gender, marital status, pay level, education, years of service in the military, finances, children, number of children in school, single parent.

Duration and location of deployment, number of children, and employment of spouse showed no significant relations.

Higher order MANOVAs (up to 3-way) showed significant relations between soldier's extent of perceived stressors and age groups and finances. Significant interactions were found for marital status by finances and for age groups by finances.

Questions for presence or absence of specific stress related symptoms (yes or no) were most often answered "no". Frequency counts showed for 6 of the 12 symptoms a relatively higher importance to the respondents. These were in rank order (number of "yes" out of 213): sleep pattern disturbances (100), changes in energy level (71), nervousness (67), anxiety (55), and weight losses or gains (54).

Correlation between soldier's extent of perceived stressors and stress related symptoms ratings was nearly zero and negative. Substantial relations between soldier's extent of perceived stressors and self ratings on coping skills could not be found either.

Discussion

The soldiers rated themselves far more often positive than negative with respect to their overall health, their stress tolerance, and stress symptoms frequency. They were only moderately affected by stressors, at least in their self-perception (see T. Hammelman, 1991). So one should be careful not to over interpret the results.

In this study the strongest influence on soldier's extent of perceived stressors was due to the financial strain imposed by ODS/S. Membership in age groups and educational level influenced the soldier's response to previous stressors, whereas factors such as location and duration of service during ODS/S as well as the total number of children the soldier had custody of did not.

The results support the following conclusion: problematic cases among mobilized reservists should mostly be expected if the individual could be subsumed in one of the following profiles:

- 1) Age up to 25, financial loss, married, lower educational level (and female)
- 2) Age 26 to 35, financial loss, married
- 3) Age 36 to 45, financial loss, married, (and male)
- 4) Age 46 and older, financial loss and higher educational level, married.
- 5) Financial loss, 3 children or more in school.

In these cases, a more in-depth exploration is recommended in order to determine the most effective support. About 41% would need some kind of financial support.

For further research, stress indicators other than only self ratings are strongly recommended as well as predeployment base-line stress levels and direct measures of stress effects in the soldier's families.

References

Hammelman, T. L. (1991) The Persian Gulf Conflict: The impact of stressors on Army Reserve Soldiers and their families. Unpublished manuscript.

Footnotes

The author wishes to thank Tracie L. Hammelman, M.S.W. who designed the questionnaire, collected the data, and gave permission to use the data for further analysis.

Current address for Ludwig Uhlmann is with the U.S. Army Health Care Studies and Clinical Investigation Activity, Health Services Command, Fort Sam Houston, Texas 78234-6060.

Army Medical Service Corps Reserve Personnel
Demobilized After Operation Desert Shield/Storm

A. David Mangelsdorff, Ph.D., M.P.H.
US Army Health Care Studies and
Clinical Investigation Activity
Health Services Command (ATTN:HSN-T)
Fort Sam Houston, Texas 78234-6060

AD-P006 962



Army Reserve medical personnel were surveyed as they demobilized after Operation Desert Shield/Storm. Responses of Medical Service Corps personnel were compared between Administrators and Health Service Providers. Predictive models were developed.

United States armed forces are structured according to a Total Force Policy. In the Army, roughly 75% of the medical assets are in the Reserve Components while the remainder are Active Components personnel. During Operation Desert Shield/Desert Storm many reserve medical unit personnel were mobilized.

Headquarters, U.S. Army Health Services Command (HQ HSC) requested assistance in the development and scoring of a questionnaire to assess attitudes of the 10,000 Army Reserve medical personnel from 53 units who were mobilized to stations in continental United States during Operation Desert Shield/Storm. The completed report details the findings (Mangelsdorff, Twist & Moses, 1991). This report will examine the results of surveys administered to Army Reserve Medical Service Corps personnel during demobilization after Desert Storm.

Method

Survey packets were sent from HQ HSC to installations where reserve units were demobilizing after Desert Storm. Surveys were administered, collected, and return to HQ HSC for processing. The surveys consisted of 13 demographic and 38 6-point Likert items (1 = strongly disagree to 5 = strongly agree, 6 = non applicable/missing). Descriptive and comparative statistics were computed.

Results

Survey responses were received from 3,930 reservists. The overall sample was 58.6% male, 15.7% dual family member, 59.4% married. Personnel were grouped according to rank, years of service, occupational specialty, and reserve category (Troop Program Unit, Individual Mobilization Augmentee, Individual Ready Reserve). The reserve personnel had an average of 10.6 years in service.

92-17695



Of the reservists responding, 244 were Medical Service Corps officers. Of these, 121 were in the 67 MOS series (Administrators) while 123 were in the 68 MOS series (Health service providers). The Administrators had an average of 16.0 years of service, 71.9% were in Troop Program Units (TPUs), 78.5% were male, 12.3% dual family member, 69.4% married, 48.7% were company grade, and 25.5% had 21 or more years of service. The Health service providers had an average of 17.1 years of service, 73.9% in TPUs, 89.4% male, 9.7% dual family member, 84.5% married, 29.2% company grade officers, and 30.0% had 21 or more years of service.

There were significant differences between the Administrators and the Health service providers on a number of items. The Health service providers were less satisfied with the clarity of the reporting instructions to the mobilization station ($p < .01$). The Administrators were less satisfied throughout the mobilization process with the communication from the installation ($p < .02$). The Health service providers believed their active duty assignment was more similar to their reserve assignment ($p < .01$). The Administrators reported having suffered financially more because of being mobilized ($p < .08$), with the effect being greatest for those who were married ($p < .05$).

The stepwise multiple regression models for predicting "probability remain in service until eligible to retire" showed differences between Administrators and Health service providers. For Administrators, the most salient variables were "At time called, unit was well prepared", "Active duty assignment is similar to reserve assignment", and "Years of service." For the Health service providers, the most salient variables were "Transition from the reserve components to active duty was easy" and "Eager to serve country."

The reservists were pleased with their experiences, though there were significant concerns expressed about the lack of communication and information provided. The fragmentation of units was not adequately explained.

When there was communication from the parent unit, there was more likely to be communication from the installation, and support from the parent unit. Reservists from units that provided the information were well prepared and felt they contributed to the mission.

The reservists were eager to serve their country. The soldier's participation was supported by the spouse. Since a high percentage of the sample were married, having family support is important. The support of the spouse was critical in soldiers planning to remain in the reserves until eligible to retire.

Soldiers who felt they were well utilized during mobilization were likely to report their contribution to the mission was significant and that they were given responsibilities commensurate with their rank and expertise. The reservists felt part of the active Army medical team at the receiving units.

Discussion

As the United States Army involves its Reserve and National Guard units more in Total Army efforts, it becomes important to assess the morale and cohesion of Reserve and Guard units. Mangelsdorff and associates (1988, 1990) have examined unit climate and morale in some Reserve and Guard units; the levels of morale of selected Reserve and Guard units were comparable to those of active duty units.

In the survey of reservists being demobilized after Operation Desert Storm, the sample was predominantly from Troop Program Units. In general, the Medical Service Corps reservists were pleased with their experiences, though there were significant concerns expressed about the lack of communication and information provided. The fragmentation of units was not adequately explained. These findings paralleled those found in the overall sample.

The Administrators felt they had suffered more financially because of being mobilized; this might be explained as they were of lower rank and had less time in service. The effect was most noticeable with the married personnel.

As unit members, reserve personnel report pride and unit cohesion. Reservists reported to have maintained their professional and military skills; the majority reported being ready if mobilized. After being mobilized, some attitudes changed. As reported during the demobilization process, there were a number of dissatisfiers particularly with respect to the fragmentation of units, redistribution of personnel, financial losses associated with mobilization, and the inadequacy of communication of information. Concerns for family support and business care plans have become more critical. These factors may affect the decisions to resign or remain in the reserves. Based on the variations between the Army Medical Service Corps Administrators and Health service providers, different factors motivate the reservists and may play a part in determining whether they remain in the reserves. Army leadership and policy makers will have to address these concerns.

References

- Mangelsdorff, A.D. (1988). Texas National Guard: Unit Climate Survey Pilot Project. (Consultation Report #88-004). Fort Sam Houston, TX: Health Care Studies and Clinical Investigation Activity.
- Mangelsdorff, A.D., Stokes, J.W., & Jacobs, H. (1990). Cohesion in Army Reserve Units. In Proceedings, Psychology in the DoD, 12th Symposium, April 1990, USAFA-TR-90-1, pp. 303-307. Colorado Springs, CO: USAF Academy Department of Behavioral Sciences and Leadership.
- Mangelsdorff, A.D., Twist, P.A., & Moses, G.R. (1991). Survey of Mobilized Reserve Components Army Medical Personnel. (Consultation Report #91-008). Fort Sam Houston, TX: Health Care Studies and Clinical Investigation Activity.

U.S. Navy Medical Reservists Post Operation Desert Storm Survey

AD-P006 963



D. Stephen Nice, Ph.D.
Naval Health Research Center
San Diego, California 92138-9174

Subsequent to the recall of 9,700 Navy medical reservists for Operations Desert Shield/Storm, the Navy Surgeon General requested an evaluation of the recall process and its potential impact on turnover in the Navy medical reserves. In June, 1991, an 82-item survey was distributed; by September 3,804 medical reservists (39%) responded. Reservists were reasonably well satisfied with their recall experience, but identified a number of issues for improvement.

Approximately two weeks after the Iraqi invasion of Kuwait, the Secretary of Defense requested that the military services develop a plan for the call-up of reserve forces in support of Operation Desert Shield. The recall was authorized by the President on 22 August 1990, and set in motion a process which would activate approximately 9,700 U.S. Navy medical reservists over the succeeding four months. Recognizing this opportunity to evaluate the Total Force concept, the Surgeon General of the navy requested the Naval Health Research Center to evaluate the recall process and its potential impact on turnover in the medical reserves.

In June, 1991, an 82-item survey was distributed to all U.S. Navy medical reservists who were recalled during Operation Desert Shield/Storm (N=9,747). This survey was constructed to assess demographic information, in-processing, out-processing, recall assignment, turnover intentions, and attitudes and perceptions regarding recall issues. In order to provide objective assessments of selected issues as well as broad coverage of attitudes and perceptions both quantitative information, using Likert-type rating scales, and qualitative input, using semi-structured narrative responses, was collected.

A total of 3,804 Navy medical reservists (39%) responded to the survey between June and September, 1991. The demographic composition of the sample was very similar to the population; it included Hospital Corpsmen (56%), Nurse Corps (24%), Medical Corps (12%), and Medical Service Corps (6%). The sample was primarily white (84%) and approximately evenly divided between men (54%) and women (46%). The mean age of the sample was 35 (range;18-65). More than half of the men (70%) and women (59%) were married, and about one-half of all respondents had children living at home. Approximately 10% of the women and 3% of the men in the sample were single parents and about 25% of the spouses were either active duty military or reservists.

While many reservists were deployed to the Persian Gulf area, most (71%) served in the continental United States (CONUS), typically "backfilling" hospital or clinic jobs vacated by active duty personnel who were deployed. For those reservists assigned in CONUS, the median distance between home and recall assignment was 225 miles, and about one-third remained within commuting distance (50 miles) of their home.

92-17696



A set of 34 items developed by subject matter experts at the Bureau of Medicine and Surgery was presented in a 5-point Likert-type format to provide a quantitative assessment of attitudes and perceptions regarding the recall experience. In Order to improve the identification and interpretability of issues and facilitate statistical analyses, factor analytic procedures were used to derive a set of underlying dimensions. Factor loadings of .40 or greater were used to develop factors, and a Varimax rotation was employed. This principal components analysis yielded a nine-factor solution. The following descriptive labels were assigned on the basis of item content: (1) satisfaction with in/out processing, (3) preparation in the reserves, (3) assignment satisfaction, (4) preparedness, (5) command staffing and equipment, (6) habitability and administration, (7) community and family support, (8) financial and family hardship, and (9) school attendance and dependent care hardship. The items in these factors were then entered into a scaling analysis to determine their internal consistency and reliability. The reliability of the school attendance and dependent care scale was not sufficient (coefficient $\alpha = .45$) and this two-item scale was removed from subsequent analyses. The remaining reliabilities (coefficient α) ranged from .69 to .89.

The overall mean levels of satisfaction on each of the 5-point scales were preparedness (4.33), community and family support (3.82), satisfaction with in/out processing (3.77), assignment satisfaction (3.61), habitability (3.44), command staffing and equipment (3.43), financial and family hardship (3.36), and preparation in the reserves (3.22). A between-group analysis of variance was computed to assess the relationship between assignment (Persian Gulf versus CONUS) and occupation (Medical Corps, Nurse Corps, Medical Service Corps, Hospital Corps) on each of the eight satisfaction scales. Generally speaking, personnel assigned to CONUS facilities were more satisfied with their level of preparedness, their assignment, the command staffing and equipment, and habitability and administration, and experienced less financial/family hardship than their peers who were assigned to the Persian Gulf area. Personnel assigned to the Persian Gulf, however, perceived a greater level of community and family support. The main effect of occupation was significant across all scales and the general finding was that the Hospital Corpsmen were least satisfied and Medical Service Corps officers were most satisfied. Medical officers, however, reported the greatest levels of financial/family hardship.

Only about 16% of the reservists in this sample intend to resign from the reserves or transfer to the individual ready reserves. The primary reasons provided for intent to leave the reserves were civilian life issues (34%), pay issues (5%), recall management (18%), duty station issues (29%), and general military issues (14%). A content analysis of the narrative responses to the survey indicated that the most positive aspects of the recall were camaraderie (45%), skills enhancement (31%), self-growth (18%), expression of patriotism (18%), military lifestyle (16%), and job contribution (15%). The most negative aspects of the recall were family separation (17%), relationship with active duty personnel (17%), leadership (17%), disbursing/processing of payment (16%), bias (14%), living conditions (12%), and poor job fit (11%). Given the opportunity for multiple responses per item, the percentages sum to more than 100.

These data indicate that reservists were reasonably well satisfied with their recall assignment and felt prepared to perform their duties; however, a number of important issues were identified for improvement. Detailed information from this survey was presented to the Surgeon General and a task force which was assembled to make policy recommendations regarding the Navy medical reservist community and the recall process. Additional efforts are underway to assess the impact of Operations Desert Shield/Storm on the Navy medical active duty community.

**"Observations on the Role of Cohesion
in Operation Desert Storm"**

Panel Members and Affiliations

**COL Frederick Manning, Chair
Director, Division of Neuropsychiatry
Walter Reed Army Institute of Research**

**LTC Robert Gifford
Kathleen M. Wright, Ph.D.
Department of Military Psychiatry
Walter Reed Army Institute of Research**

**CPT(P) Paul Bartone
Department of Military Psychiatry
Walter Reed Army Institute of Research**

**James Griffith, Ph.D.
Westat, Inc.
MAJ Carol Nicholson
Office of the Chief, Army Reserve**

**Beverly C. Harris, Ph.D.
Alma G. Steinberg, Ph.D.
U.S. Army Research Institute**

Abstract

The panel consists of four papers which examine the construct characteristics of cohesion and their relationship to specific antecedent and consequent conditions. Findings draw from several ongoing data collection efforts seeking to better understand the stresses of Operation Desert Storm (ODS) and the role of human factors in buffering these stresses on the individual soldier and on improving combat performance. Specifically, the papers will present:

- (1) Observations on physical, psychological, and wartime stresses in ODS and the extent to which cohesion prevented de-habilitation among active duty Army personnel;
- (2) Survey findings on the relationship of vertical and horizontal cohesion to the incidence of combat stress reactions among active duty Army personnel participating in ODS;
- (3) Survey results on the perceptions of deployment experiences, quality of unit leadership, cohesion, and team training found in Army Reserve re-constituted (or "cross-leveled") units and in intact units; and
- (4) Survey results on the unique experiences of and level of cohesion and morale found among individually deployed reservists (soldiers in the Individual Ready Reserve), compared to those found among deployed Army National Guard and Army Reserve soldiers.

The discussion section of the panel will focus on: (1) a summary of findings and what they mean for developing the construct of cohesion; (2) a statement of the role of cohesion in ODS and in future wars; and (3) the preeminence of cohesion and other human factors relating to combat performance for future military research.

**Cohesion and Post-Combat Adjustment Following
Operation Desert Storm**

CPT (P) Paul T. Bartone
Department of Psychology
Walter Reed Army Institute of Research
Washington, DC

Operation Desert Storm exposed U.S. soldiers to a variety of combat stressors. Historically, small unit cohesion has been identified as a "combat multiplier" in that it can enhance performance by facilitating teamwork, motivation, and commitment to the mission. Cohesion has also been shown to increase the "staying power" of soldiers, with fewer desertions and psychiatric casualties in cohesive units.

The present study examines the influence of cohesion on a variety of indices of psychological adjustment following Operation Desert Storm. A research team led by Dr. David H. Marlowe, Chief of the Department of Military Psychiatry, Walter Reed Army Institute of Research (WRAIR), collected post combat "debriefing" surveys on 5,391 U.S. Army soldiers who deployed to Saudi Arabia during the Persian Gulf War. All surveys were administered 4-8 months following the end of the War.

Multiple regression analyses confirmed the contribution of both vertical and horizontal cohesion to healthy psychological adjustment following combat exposure.

**Building Cohesion in the Desert:
Observations of Army Units in Operation Desert Shield**

**LTC Robert K. Gifford
Kathleen M. Wright, Ph.D.
Department of Military Psychiatry
Walter Reed Army Institute of Research
Washington, DC**

As the U.S. Army prepared for war in the Persian Gulf, many units were challenged by the need to integrate new personnel, establish new intra-unit relationships as they formed combined arms teams and task forces through cross-attachment of sub-units, and form new working relationships with other units. these tasks conditions, into a harsh desert environment in a cultural setting unfamiliar to most Americans. The need to prepare simultaneously for diverse missions, including defense, deterrence, and offensive operations, added complexity to the situation.

Soldiers' reactions to such issues were addressed in interviews with forward-deployed units in Saudi Arabia, conducted by research teams under the direction of Dr. David H. Marlowe, Chief, Department of Military Psychiatry, WRAIR, during Operation Desert Shield. Observations from those interviews are presented to show the social and psychological processes operating as units coalesced into a combat-ready force.

**"Cross-Leveling" of U.S. Army Reserve Soldiers
During Operation Desert Storm
and Its Relationship to Cohesion**

James Griffith, Ph.D.
Westat, Inc.
Rockville, MD

LTC Joe Kulbok
Office of the Chief, Army Reserve

It has been argued that a strength of the Army reserve component (both the Army Reserve and Army National Guard) lies in its higher level of cohesion and morale, compared to its active duty Army counterpart. This assumption is based on personnel stability found in these units, in addition to these soldiers coming from the same communities. The purpose of this paper is to examine the validity of this assumption, by presenting findings on cohesion and related issues for a group of Army Reservists who were deployed with and without their units.

Each year, a probability survey of Army Reservists is conducted. This year, the design included a probability sample of Reservists who were deployed in support of Operation Desert Storm and those who were not deployed. Among the deployed, soldiers were deployed with their units ($N = 3,566$) and without their units ($N = 546$). The paper presents differences between these two groups of deployed soldiers, focusing on differences in cohesion and relating these differences to combat readiness. Findings are discussed in terms of the benefits of unit versus individual replacement systems.

Issues Related to the Motivation and Cohesion of Mobilized Reserve Soldiers

Beverly C. Harris, Ph.D.
Alma G. Steinberg, Ph.D.
U.S. Army Research Institute
Alexandria, VA

This paper addresses factors relating to the motivation and cohesion of U.S. Army Reserve soldiers who were mobilized for Operation Desert Storm (ODS). Data were obtained from a total of 3,668 Individual Ready Reserve (IRR), U.S. Army Reserve (USAR), and Army National Guard (ARNG) soldiers activated for ODS. The IRR consists of individuals who have left the Army to return to civilian life but are still obligated to military service. They remain on the Army roster until their obligation is completed. However, they do not train or drill during this time nor are they members of units. Soldiers in the Troop Program Units of the USAR and ARNG, on the other hand, are part of the Selected Reserve, and train and drill monthly in units.

This paper presents survey findings on attitudes and concerns at call-up of these different Army reserve populations during, and after their mobilization and deployment for ODS. Topics covered include their experiences during call-up; their levels of motivation, cohesion, and unit morale; and their perspectives with respect to their peers, leaders, and the Army as a whole (horizontal, vertical, and organizational bonding). Implications for increasing motivation and cohesion of Army reserve component soldiers in future mobilizations are discussed.

**Army Family Stress, Support, and Coping
in Operation Desert Shield/Storm**

Joel M. Teitelbaum, Ph.D., Walter Reed Army Institute
of Research [WRAIR], Panel Organizer
Sally Bell, U.S. Army Research Institute [ARI]
Leora N. Rosen, Ph.D., WRAIR
Florence Rosenberg, Ph.D., WRAIR & USUHS
CH (MAJ) Maria J. Snyder, U.S. Army Chaplaincy
Services Support Agency [USACSSA]
CH (LTC) John A. Well, USACSSA
LTC David J. Westhuis, U.S. Army Health Care Studies
and Clinical Investigation Activity [HCSCIA]

Abstract

This Panel consists of multi-disciplinary members from the U.S. Army's Operation Desert Shield/Storm (ODS) Family Factors Research Task Force. These researchers performed a social-psychological field study of family members of deployed soldiers using sample survey questionnaires and observation/interview methods. The study assesses ODS stressors and the stress perceptions of waiting spouses, formal and informal social supports, and their effects on spouse coping and adaptation to wartime separation and soldier return from ODS. The study was performed during the overseas mission on cluster samples of spouses in targeted deployed Active combat and support units and Reserve (ARNG/USAR) support units and their Army support givers.

The Task Force addressed the effects of sub-acute stressors from a prolonged, uncertain, and hazardous military deployment to the Middle East for up to nine months, August 1990 to May 1991. The panel will discuss leading military and family life stressors and stress perceptions among military spouses. It will present information on key Army stress prevention and stress mediating institutions, including: volunteer Family Support Groups (FSGs) composed of unit spouses, other family members and soldiers; unit-level Rear Detachment Commands (RDCs) which cared for families during the soldiers' absence; installation, division, and Army community service provider programs and Family Assistance Centers (FACs). The short and longer term effects of deployment separation on marital stability, divorce, and family functioning during and after ODS will be examined.

The panel encourages interactive audience responses and comments on ODS lessons learned and on organizational factors in generalizable to family stressors, support and coping in major combat missions for all armed services. Discussion will include the effects of spouse perceptions on family and child well being and adaptation during wartime deployments, and the impact of the family-military interface on individual soldier psychological readiness and deployability, and unit cohesion during hazardous combat missions.

Author/Presenter Index

- Arends, 226
Aretz, 138, 183
Ashley, 91
Austin, 249
Baisden, 217, 227
Baker, 133
Ballard, 128, 228
Barrow, 255
Bartone, 281, 283
Baruch, 11
Becker, 163
Belenky, 283
Bell, 287
Bensel, 254
Berger, 6
Bergondy, 133
Beringer, 123
Blake, 229
Blower, 66
Bryant, 109
Calhoun, 96
Carlgren, 230
Carney, 257
Coccia, 231
Conroy, 113
Cooney, 230
Crane, 148
Cuervo, 232
Cummings, 104
Curphy, 96, 257, 258, 263
Dansby, 250, 264
Davies, 255
Davis, 233
Degenhardt, 237
Dilla, 36
Dolgin, 66
Engberson, 234
Entin, 193
Faley, 250
Fentress, 235
Gajeski, 128
Garcia, 254
Gibb, 236
Gibson, F., 91, 96, 257, 258, 263
Gibson, G., 202
Gifford, 237, 281, 283
Gould, 238
Griffith, 268, 281
Grisham, 56
Grunzke, 113
Haney, 236
Harris, 281
Hendrix, 202
Higgins, 249
Hogan, J., 257, 263
Hogan, T., 257, 263
Holcombs, 239
Holgate, 254
Horn, 113, 163, 168,
Houston, 257
Howes, 245
Huang, 255
Hubbard, 11
Hughes, J., 173
Hughes, R., 263
Humphrey, 158
Idar, 240
Irish, 250, 264
Jackson, 1
Joe, 264
Jones, G., 207
Jones, S., 254
Jones, W., 168, 241
Jordan, 242
Kain, 228
Keller, 163, 237
Kirkpatrick, 163
Knight, 51
Kramer, 158
Kulbok, 281
Lamb, 198
Landis, 250, 264
Landreth, 258
Landry, 153
Lautenschlager, 243
Leber, 173
Leitham, 244
Lindholm, 178
Lindsay, 91
Macomber, 96

Author/Presenter Index

- Mallery, 6
Mangelsdorff, 268
Manning, 281
Marlowe, 283
Martin, E., 178
Martin, J., 221
Martin, M., 207
Mastroianni, 6
Matthews, M., 100, 212
Matthews, V., 242
McCloy, 21
McCown, 257, 258
McDowell, 163
Micalizzi, 16
Mitchell, 51
Mueller, 86
Neibuhr, 250
Nice, 268
Nicholson, 281
Nilsen, 263
Nontasak, 66
Olive, 228
Orrell, 163, 243
Oswald, 250
Parker, 61
Patterson, 243
Peeples, 168
Perry, 268
Peterson, 81
Pine, 230
Porter, 16, 26, 41, 61
Potter, 229
Prince, 133
Prue, 245
Rainey, 46
Rate, 183
Riggs, 71
Robinson, 245
Roland, 246
Rosebush, 31
Rosen, 287
Rosenbach, 86
Rosenberg, 287
Rosenfield, 264
Salas, 133
Sashkin, 86
Scaramozzino, 240
Schoen, 6
Seidler, 118
Serfoss, 11
Shane, 228
Slimack, 229
Smith, 236
Snyder, 287
Stanny, 158
Steinberg, 281
Stout, 133
Street, 66
Strevell, 163
Stricker, 21
Swindells, 236
Tallarigo, 247, 248
Teitelbaum, 287
Thurber, 61
Travis, 66, 158
Tree, 249
Uhlman, 268
Van Fleet, 81
Vance, 188
Viera, 235
Walker, 173
Warner, 11
Weaver, 100
Wegner, 76
Wells, 287
West, 36
Westhuis, 287
White, 244
Wickens, 113, 118, 143
Wittman, 6
Worringham, 123
Worthington, 198
Wright, K., 281, 283
Yoos, 46
Ziebell, 71