

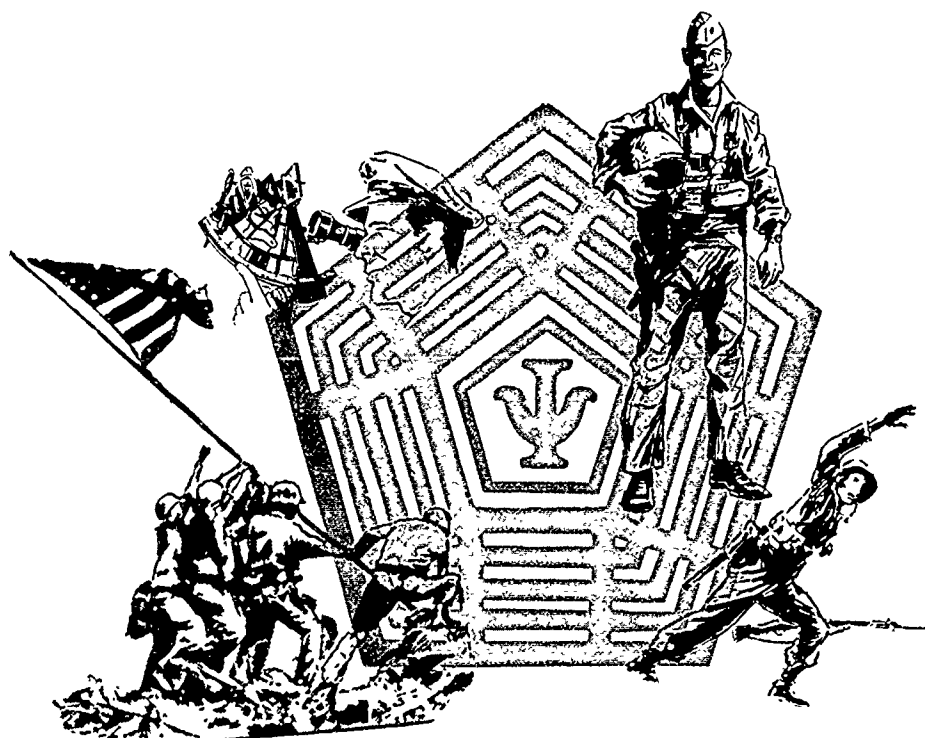
AD-A240 618



USAFA-TR-82-10

1

PSYCHOLOGY IN THE DEPARTMENT OF DEFENSE



DTIC
ELECTE
SEP 25 1991
S D D

EIGHTH SYMPOSIUM
21-23 APRIL 1982

SPONSOR:

USAF ACADEMY

DEPARTMENT OF BEHAVIORAL SCIENCES AND LEADERSHIP

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

01924 050

91-11395



PROCEEDINGS OF THE EIGHTH SYMPOSIUM
PSYCHOLOGY IN THE DEPARTMENT OF DEFENSE
21 - 23 April 1982

Department of Behavioral Sciences and Leadership
United States Air Force Academy
Colorado Springs, Colorado

Symposium Chairman
Major Gail I. Arnott

SYMPOSIUM COMMITTEES

Professor and Head, Department of
Behavioral Sciences and Leadership

Colonel John W. Williams, Jr.

SYMPOSIUM CHAIR

Major Gail I. Arnott

EXECUTIVE COMMITTEE

Major Gail I. Arnott, Chair
Colonel John W. Williams, Jr.
Colonel William E. Rosenbach
Major Richard L. Hughes
Major John F. Swiney, Jr.
Major Thomas M. McCloy
Major Robert A. Gregory

PROGRAM COMMITTEE

Major John F. Swiney, Jr., Chair
Lieutenant Colonel John E. Anderson
Major William L. Derrick
Major Charles D. Gorman
Captain Mickey R. Dansby
Captain William H. Cummings III

INVITATIONS AND ATTENDANCE COMMITTEE

Major Mark Nataupsky, Chair
Lieutenant Colonel Valentin W. Tirman, Jr.
Captain William P. Marshak
Captain Philip A. Irish III

SOCIAL, BILLETING & TRANSPORTATION COMMITTEE

Major Stephen J. Pacheco, Chair
Major Richard T. Smith

FACILITIES AND VISUAL AIDS COMMITTEE

Captain Harold E. Roper, Chair
Master Sergeant Frank C. Derry
Sergeant Larry Dannelley

PROTOCOL COMMITTEE

Major Robert C. Ginnett, Chair
Captain Melinda Mitchell
Captain Ruth E. Freese

PUBLICITY COMMITTEE

Captain Linda D. Jackson, Chair
Major Joseph W. Evans, Jr.

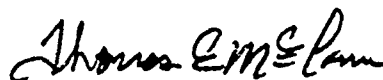


Accession For	
NTIS	CRASH
DTIC	TAB
Unannounced	
Justification	
By	
Distribution	
Availability Codes	
Dist	Availability for Special
A-1	

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

This report has been cleared for open publication and/or public release by the appropriate Office of Information in accordance with AFR 190-1/ and DODD 5230.0. There is no objection to unlimited distribution of this report to the public at large, or by DDC to the National Technical Information Service.

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



THOMAS E. McCANN, Lt Colonel, USAF
Director of Research and
Continuing Education

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

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER USAFA TR 82-10	2. GOVT ACCESSION NO. ADA-	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Proceedings, Psychology in the Department of Defense, Eighth Annual Symposium		5. TYPE OF REPORT & PERIOD COVERED Proceedings
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Major John F. Swiney, Jr. (Ed.)		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Behavioral Sciences and Leadership U. S. Air Force Academy Colorado Springs, CO 80840		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS HQ USAFA/DFBL USAF Academy, Colorado Springs, CO 80840		12. REPORT DATE December 1982
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for Public Release and Distribution Unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Clinical Psychology, Human Factors, Measurement & Evaluation, Military Family, Organizational Development, Performance Evaluation, Personality Assessment, Personnel, Pilot Performance, Psychology, Radiation Biology, Relocation Stress, Social Support Systems, Stress Management, Symposium Training, Training Research Evaluation, Turnover Decision, Visual Display & Processing.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) These printed proceedings include papers and presentations that deal with a wide range of research in psychology with emphasis on military issues. 25 * Psychology, * Military Research, * Summary		

FORWARD

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

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

These proceedings are approved for public release, distribution unlimited, in accordance with paragraph 1VA, AFR 80-45.

In publishing the Proceedings of the Eighth 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

Editorial Assistance: Helen Wilson and Rick Conner

OVERVIEW OF SYMPOSIUM SCHEDULE

Wednesday, 21 April 1982

Morning Activities

Event

Registration

Keynote Address

Dr. Larry L. Cummings

Panel Discussion and Paper Sessions

Panel: The Theory and Practice of Coping
with Relocation Stress

Paper: Visual Performance

Paper: Personnel Evaluation

Invited Luncheon Address

Colonel James E. Baker

Afternoon Activities

Paper Sessions and Panel Discussions

Paper: Human Factors

Paper: Training and Training Research

Paper: Social Psychology

Panel: Military Family Issues and Programs
in the 1980's

Panel: Team Training and Performance: Research
for the 80's

Paper: Stress Management

Thursday, 22 April 1982

Morning Activities

Panel Discussion and Paper Session

Panel: Desirability of the Alternatives: A
Promising Construct for Under-
standing Turnover Decisions

(22 Apr 82 Continued)

Paper: The Military Family

Invited Address
Dr. Fred E. Fiedler

Symposium Luncheon Address
Dr. Sharon B. Lord

Afternoon Activities

Paper Sessions, Presentation, Tutorial,
and Panel Discussion

Paper: Measurement and Evaluation

Presentation: Psychologists in the Air
Force: Utilization, Structure
and Job Attitudes

Paper: Clinical Psychology

Tutorial: Leading Skills for Situational
Leadership

Panel: Management of Separations and
Discharges--Work in Progress

Paper: Personnel Issues

Paper: Values and Social Systems in the
Military

Paper: Human Factors II

Friday, 23 April 1982

Morning Activities

Panel Discussions, Paper Sessions, and an
Invited Address

Panel: Personality Assessment in Operational
Settings

Paper: Physiological Correlates of Stress

Paper: Performance Evaluation

Invited Address
Dr. Charles B. Spielberger

Panel: Soldier-System Interaction: ARI's
Systems Research Laboratory

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

WEDNESDAY, 21 APRIL 1982

Page

KEYNOTE ADDRESS

Dr. Larry L. Cummings

PANEL DISCUSSION: The Theory and Practice of Coping with Relocation Stress

Chair: Ralph R. Nielsen

Panelists: Patricia C. Nida

Peter M. Hayes

Carol Hayes

Sharon Nielsen

PAPER SESSION: Visual Performance

Chair: William L. Derrick

Interactions Between Pattern and Motion in the Visual Perception of
Moving Objects

William P. Marshak, Chester D. Wilson, and Rand L. Brown 1

Proposed New Vision Standards: Contrast Sensitivity

David W. Evans and Arthur P. Ginsburg 13

Pilot Selection for Target Detection

Russell A. Benel and Denise C.R. Benel 22

Effects of Task Performance on Visual Accommodation and Perceived Size

Valerie J. Gawron and Kenneth R. Paap 27

Visual Cues in the Simulation of Low Level Flight

Edward J. Rinalducci, Elizabeth C. Martin and Thomas Longridge 32

PAPER SESSION: Personnel Evaluation

Chair: Thomas E. Ulrich

Selection for Command: What Really Counts?

Frederick V. Malmstrom, Edward Batchelor, Jr., Christopher J. Reamy, and
Richard M. Coffman 37

Development and Implementation of the Officer Training School Selection
Equation

Lynn M. Scott 42

Nontraditional Admissions Factors: The Special 100 Project

Harald E. Jensen and William J. Strickland 48

Characteristics of U.S. Navy Recruits on Reading Comprehension and Selected Educationally Related Variables Norman J. Kerr and Harry L. Bowman	57
Validation of the Computer Programmer Aptitude Battery J. P. McMenemy	63
PAPER SESSION: Human Factors Chair: Thomas M. McCloy	
Vibration Effects on Digit-Symbol Coding Alvah C. Bittner, Jr., John C. Guignard, Jeffrey C. Wolstad, and Robert C. Carter	69
The Celestial Training Device - An Example of a Low Fidelity, Highly Effective Part Task Trainer Ennis R. Pennington	77
Use of Optical and Thermal Sights in Daylight Target Detection Brian L. Kottas and David W. Bessemer	83
PAPER SESSION: Training and Training Research Chair: William H. Cummings III	
Reverse-Role Training K. Ronald Laughery, Jr. and Jan L. Ditzian	88
The Effects of Human Relations Training on the Leadership Qualities of Military Personnel William M. Barkley and Richard L. Percy	97
Use of Weaponeer Marksmanship Trainer in Predicting M16A1 Qualification Performance Joel D. Schendel and Frederick H. Heller	106
Development of Basic Skills Instruction for the Army: Analysis and Verification of Basic Skill Requirements for One Military Occupational Specialty Eva D. Vaughan and Zita Glasgow	111
Equivalency of Simulator Exercises to Real World Experience in Meeting IMCO Training Standards Kent E. Williams and Phyllis J. Kayten	118
Training Program Evaluation: An Application of Discrepancy Evaluation to Military Training Donald M. Kristiansen	132
Assessment of Training James L. Sherrill	137

PAPER SESSION: Social Psychology

Chair: Mickey R. Dansby

Relation of Cooperative Response to Interpersonal Needs Jarean Ray and Paul Brown	150
Effects of "Hyper" Descriptor and Cadets Second Order Self Perception on Their Ratings of a Hypothetical Cadet David B. Porter and June I.R. Babson	158
Nonverbal Communication in Military Counseling Gary A. Packard, Jr., and Mickey R. Dansby	166
Stopout Returnees to USAFA Jim Clifford	170
The Interdependence of International Affairs: Social Psychology as a Politico-Military Tool Richard W. Bloom	176
PANEL DISCUSSION: Military Family Issues and Programs in the 1980's	
Chair: James E. Cochran	
Panelists: Ann O'Keefe	
Betty Hart	
John F. Gimber	
Lyle A. Working	
Discussant: Joseph W. Kovarick	
PANEL DISCUSSION: Team Training and Performance: Research for the 1980's	
Chair: Jean L. Dyer	
Panelists: Bertram Cream	
Lawrence Reed	
Larry Nadler	
Rohn J. Hritz	
Thomas J. Roth	
Donald W. McGill	
Team Training Research: Into the Unknown? Jean L. Dyer	182
A Taxonomic Model for Navy Teams Larry Nadler	187
A Method for Describing Teams and Team Behavior Rohn J. Hritz, Thomas J. Roth, Donald W. McGill and Jean L. Dyer	193
An Empirically Derived Descriptive Model of Team Behavior: The Concept of Dependency Rohn J. Hritz, J. Thomas Roth, Donald W. McGill, and Jean L. Dyer	198

PAPER SESSION: Stress Management

Chair: Robert C. Ginnett

Stress Management: A Model for Optimizing Human Resource Effectiveness T. Roger Manley	203
Battlefield Stress: Management Techniques Russell J. Hibler	209
What Experience Teaches: Cognitive Responses to a Stressful Environment Earl H. Potter III	219
Training for Combat Stress Reactions A. David Mangelsdorff and T. Paul Furukawa	228
Effect of Individual and Organizational Variables on Perceived Job Stress William H. Hendrix, Nestor K. Ovalle, Raymond G. Troxler, Charles W. Staton and Samuel P. Fye	232
A Pilot Evaluation of Stress Management Techniques for Collegiate Rifle Shooters Kenneth C. Coon and Richard L. Hughes	237

THURSDAY, 22 APRIL 1982

PANEL DISCUSSION: Desirability of the Alternatives: A Promising Construct
for Understanding Turnover Decisions

Chair: Victor H. Appel

Panelists: Thomas W. Watson
Gerald D. McCright
James D. Warren

The Desirability of Alternatives: A Promising Construct for Understanding Turnover Decisions Thomas W. Watson and Victor H. Appel	242
Examining the Turnover Processes Within the Framework of an "Affect Mini- mization" Paradigm Gerald D. McCright	247
The Greener Grass Phenomenon: Perceptual Distortion in Individual Decision-Making James D. Warren and Victor H. Appel	252
PAPER SESSION: The Military Family	
Chair: Harold E. Roper	
Air Force Wives and Officers Wives Clubs Marianne Pierce and Vince Luchsinger	260

	Page
Air Force Family Support Centers Bobby Donohue and Vince P. Luchsinger	265
Job Satisfaction Among U.S. Air Force Academy Homemakers David B. Porter and Sharon J. Porter	270
Dual-Career Family Negotiations Among USAF Academy Cadets Frank R. Wood and Leona Flores	277
INVITED ADDRESS: The Contribution of Intellectual Ability and Organizational Experience to Leadership Performance Fred E. Fiedler	283
SYMPOSIUM LUNCHEON ADDRESS Sharon B. Lord	298
PAPER SESSION: Measurement and Evaluation Chair: John F. Swiney, Jr.	
Structures of Memory for Critical Flight Information Roger W. Schvaneveldt, Timothy E. Goldsmith and Francis T. Durso	304
A Video Game as a Covariate for Carrier Landing Research Gavan Lintern and Robert S. Kennedy	311
Standard Errors Associated with Item Response Theory Parameters Benjamin A. Fairbank, Jr., David Thissen, Howard Wainer, Michael Levine and Malcolm Ree	317
Averaged Correlations Between Parallel Measures: Reliability Estimation Alvah C. Bittner, Jr.	321
Development of the Common Metric Bennie W. Roach and Deborah L. Rogers	328
The Use of Conjoint Analysis in the Interval Subjective Scaling of Mental Workload Stephen P. Boyd	334
SINGLE PRESENTATION: Psychologists in the Air Force: Utilization, Structure, and Job Attitudes Jimmy L. Mitchell, James B. Keeth, and Linda A. Wiekhorst	342
PAPER SESSION: Clinical Psychology Chair: John E. Anderson	
The Ability to Say 'No' as a Mental Health Factor Allan H. Frankle	352
The AFMET Project and the Bloom Sentence Completion Surveys Wallace Bloom	362

Human Development Center - A Movement Toward Preventive Mental Health David A. Giles	368
The Interpersonal Dimensions of the Passive-Dependent and Compulsive Personalities in the Military Frank E. Emmett, Jr.	371
Rational-Emotive Disputing Technique: An Approach to Reduce Prejudice Z. Benjamin Blanding	383
TUTORIAL: Leading Skills for Situational Leadership *Loren L. Moore	
PANEL DISCUSSION: Management of Separations and Discharges--Work in Progress Chair: Barry E. Goodstadt	
Panelists: William H. Githens Ann Majchrzak Mary Beth Merrin Glenda Y. Nogami Nora Kinzer	
Discussant: H. Wallace Sinaiko	
Management Actions to Reduce UA and Non-EAS Attrition in the U.S. Marine Corps Ann Majchrzak	392
The Impact of Management Variables on Attrition in Units Mary Beth Merrin	399
An Empirical Investigation of First Term Attrition Glenda Y. Nogami	403
Enlisted Attrition: Process and Outcome Glenda Y. Nogami	410
PAPER SESSION: Personnel Issues Chair: Charles D. Gorman	
A Preliminary Look at Vocational Interest and Attrition Among Air Force Enlistees Michael D. Matthews and Gene A. Berry	415
Personality Characteristics of Entering Air Force Academy Cadets Richard L. Hughes	420
Determining Officer Education Requirements Lynn M. Scott	425

* Manuscript not available for publication

	Page
A Hybrid Organizational Development Strategy for Small Group Interventions Robert C. Ginnett	431
PAPER SESSION: Values and Social Systems in the Military Chair: Melinda Mitchell	
Influential Associations in Organizations: The 'Touchstones' of Career Success Frank R. Wood and Rosanna Hertz	440
Values and the Foreign Military Officer Lynnville E. Taber	446
A Comparative Study of Values of Military Men and Women Jose Bolton	454
Analyzing the Army COHORT Experiment: Some Critical Questions Robert P. Kane	462
Hostage Psychology: The Stockholm Syndrome, Transference, and Identification With the Aggressor Victor M. Solomon	468
PAPER SESSION: Human Factors II Chair: James A. Boyless	
A Processing Resource Explanation of Subjective Dimensions of Operator Workload William L. Derrick	473
Individual Definitions of the Term "Workload" Sandra G. Hart, Mary E. Childress and Jan R. Hauser	478
The Essence of Pilot Flying Skill Edward E. Eddowes	486
The 1981 Naval and Marine Corps Aviation Anthropometry Survey and Applications Harv G. Gregoire	490
Designing for Compatibility Thomas M. McCloy, James A. Boyless and Bruce R. Jaeger	494

FRIDAY, 23 APRIL 1982

PANEL DISCUSSION: Personality Assessment in Operational Settings	
Chair: Thomas P. O'Hearn	
Panelists: E. Roger Williams	
Thomas Booth	
Ron Chapman	
Wallace Bloom	
Frank E. Emmett	

Discussant: Charles D. Spielberger

PAPER SESSION: Physiological Correlates of Stress

Chair: Colleen Turner

Atropine Fails to Counter Radiogenic Behavioral Deficits

G. Andrew Mickley, Karen E. Stevens, Gerald H. Lawrence, Gregory L.
Gibbs and Gerald A. White 498

Neutron Irradiation Produces Behavioral Changes Which Are Not
Reversed by Atropine

Karen E. Stevens, G. Andrew Mickley, Gerald H. Lawrence, Ronald G. Frass
and Robert Plonsky 507

Causal Analysis of Military Cold Injuries

James B. Sampson and Jared B. Jobe 516

Longitudinal Health Effects Associated with Aviation

Anne Hoiberg 523

PAPER SESSION: Performance Evaluation

Chair: Joseph W. Evans, Jr.

Maintenance Performance Research in the Army

John F. Hayes 527

Firefighting Performance of Navy Male and Female Shipboard Crewmembers

Ross L. Pepper and Mark D. Phillips 533

The Development of the Air Force Civilian Job Performance Appraisal System

John A. Guerrieri 540

The Air Force Job Performance Appraisal System

John A. Guerrieri 545

INVITED ADDRESS: Effects of Stress on the Performance of Military
Personnel

Charles D. Spielberger 550

PANEL SESSION: Soldier-System Interaction: ARI's Systems Research
Laboratory

Chair: Irving Alderman

Panelists: William Bickley
Jay Coke
John Lockhart
Richard Palmer
Paul Rossmeissl
Steven Steward

Aviator-Aircraft Integration: ARI Research in Army Aviation

W. R. Bickley, W. R. Brown, J. A. Dohme, and J. H. McCracken 551

INTERACTIONS BETWEEN PATTERN AND MOTION IN VISUAL PERCEPTION OF MOVING OBJECTS

by

Captain William P. Marshak
Cadet 2nd Class Chester D. Wilson
Cadet 2nd Class Rand L. Brown

Department of Behavioral Sciences and Leadership
U.S. Air Force Academy, CO 80840

INTRODUCTION

Moving stimuli presented to the human visual system elicit responses from a wide variety of visual neurons. Cells sensitive to motion or temporal modulation are activated. So are cells sensitive to the moving stimuli's contours and distribution of spatial energy. However, many cells in the visual system exhibit more than one sensitivity. They vary their firing rates with changes in both motion and pattern attributes of stimuli which enter their receptive fields. Such concurrent sensitivities strongly suggest that perceptual interactions between motion and pattern should occur in the perception of objects in motion.

One theory of visual perception has already incorporated pattern and motion interactions in its structure. Ullman's (1979) computational theory of motion perception makes explicit predictions that pattern and motion will interact. Two fundamental interpretive processes are proposed to operate on the the spatio-temporal information received from a moving stimulus. The first process is the perception of spatial structure from motion information. Examples of this are Sekuler and Levinson's (1977) phantom contours. This is a perceptual filling phenomenon that occurs when a moving stimulus surrounds a blank area; the blank area is seen to contain an illusory stimuli similar to the surrounding stimuli. Another example of pattern from motion is the apparent contour created when moving dot patterns are electronically restricted in their passage in a systematic fashion. The orderly interruption of the dot's straight line procession across the screen results in vivid apparent contours.

Ullman is not explicit about the second process, motion derived from pattern information. Neither have there been substantiated reports of this phenomenon. Although the creation of motion from pattern may eventually be observed, an easier feat might be to observe pattern influences on motion perception. The particular phenomenon that this research seeks to find an influence of stimulus pattern on perceived direction of a moving object.

An alternative theory of motion perception has been proposed by Mather (1980) and in more explicit terms by Marshak (1981). The latter theory attempts to tie together physiology and psychophysics into a comprehensive theory of how motion is perceived.

The new model is called the vector model, for reasons which will be apparent later. The vector model for motion perception consists of three stages of processing. These stages are represented as operating in a serial fashion, but this is not a necessary feature of the model.

The model's first stage consists of a collection of directionally tuned, motion sensitive units. The response to motion of these units will be based on operating characteristics of motion sensitive neurons and psychophysical evidence. Maximum activation of each unit is achieved by motion in that unit's preferred direction; less activity results from motion in directions other than its preferred direction. The directional range of activity spans 45 degrees either side of the unit's preferred direction, gradually diminishing with increased deviation from that preferred direction. A Gaussian distribution of response amplitude over directions is assumed. This and two other assumptions are used: 1) all directionally sensitive units have the same tuning functions and 2) preferred directions are equally distributed over the possible angles. The assumed Gaussian distribution of activity, homogeneity of response functions and homogeneous density of sensitive directions, are also not necessary conditions in the vector model. They are made to simplify explanation of the model.

Each DS unit's output can be represented by a vector in a polar coordinate system; hence the model's name. The orientation of the unit's vector lies along that unit's preferred direction of motion. The vector's length is proportionate to the activation of that unit. Direction of motion is the most important determinant of vector length, but, using visual neurons as a model for the unit vector, other factors may also determine vector length.

Before the broadly distributed activity among directionally sensitive units produced by a moving stimuli can serve as the basis for direction perception, some narrowing of the activity must take place. This is accomplished in the vector model by a second, inhibitory stage of processing. Upon its activation, each directionally sensitive unit exerts an inhibitory influence on units of similar preferred directions. The size of the inhibitory effect increases with directional differences from the unit's preferred direction up to a maximum, then diminishes with greater directional separation. This presumed distribution of inhibition was derived from interactions of moving directions (Marshak and Sekuler, 1979).

The final stage in direction perception is the decision stage which uses the net distribution of activation, the combination of activation and inhibition, among the DS units to determine the perceived direction.

There are two possibilities for the decision mechanism. The first mode is a peak detector. This decision mechanism would seek the longest DS unit vector and take that unit's preferred direction as the perceived direction. This is a form of the "labeled line" model used in vision and other domains (Lettvin et al., 1961). The peak detection mechanism is simple and capable of accounting for at least part of the known phenomenon.

The alternative decision mechanism is modeled after the mathematical operation of vector summation. In this more complex process, the directionally sensitive unit vectors would be summed by vector algebra to determine the perceived direction of motion. The vector sum mechanism under most conditions makes the same predictions about perceived direction as the peak detector. The summation of vectors has greater flexibility to describe other motion phenomenon, particularly motion aftereffect (Marshak, 1981). The model and its three stages are summarized in Figure 1.

Insert Figure 1 here

A less detailed but similar model was proposed by Mather (1980) as a "distribution shift" model. In his theory, Mather said there exists an array of motion sensitive elements which serve as the basis of perception of direction. Various illusions and distortions in direction perception are a result of changes in the distribution of activity within the array of motion sensitive elements. Although the distribution shift and vector models have similar underlying logic, the vector model is much more explicit about how the mechanism operates and because of this, can generate more testable predictions.

The source of pattern influence on the perception of motion's direction comes from the stage one of the vector model. Unit's vector length is determined not only by the direction of motion, but by the temporal frequencies generated by the moving object. Those temporal frequencies are a product of velocity, and the spatial frequency of the moving object. The velocity and the pattern of a moving object may produce activation of units in the vector model which have peak activation or asymmetries in activation which indicate other than the real direction of motion. This is especially true if a vector sum is used as a basis for the decision.

There is some neurophysiological data to support the notion that pattern may influence the activity of directionally sensitive cells. Movshon (1980) has recorded the directional tuning function of complex cells in cat visual cortex. When stimulated with random dot patterns moving in a single direction, cells responded to a broader range of directions than with gratings as the stimulus. Movshon attributed this broadened tuning function to the spatio-temporal characteristics of the random dot patterns. Pantle (1974) psychophysically demonstrated that the visual system responds to temporal frequencies of moving visual stimuli. Since dot patterns have power in all orientations, Movshon argued that a range of temporal frequencies are generated at varying angular differences from the direction of motion. Moving dot patterns generate temporal frequencies in cells which are sensitive to directions other than the direction of stimulus motion.

When Movshon moved dots at velocities higher than the cell's optimum velocity, a strange bi-lobed tuning function was observed. This was again attributed to the distribution of temporal frequencies over directions other than the direction in which the dots moved. At such high velocities, the resulting temporal frequencies were beyond the sensitive range of cells sensitive to the direction of motion. Frequencies are generated in other directions which fall off with a cosine function with increasing directional difference. These lower temporal frequencies are within the complex cell sensitivity range. Since those temporal frequencies can be found to either side of the direction of motion, a bi-lobed tuning function resulted.

Using Movshon's findings as a basis for the vector model, if moving stimuli were created with distinctive two dimensional Fourier spectra and were moved at certain velocities, then unit vectors in the direction of motion may not be the most active. This deviation in the distribution of unit vector activity should result in systematic misperception of the stimuli's direction of motion.

The present experiment will test this one prediction of the vector model. When an object with a non-uniform two dimensional fourier transform moves through the visual space and no background references are available, systematic errors in the perceived direction of motion should occur. These errors in perceived direction should occur in directions toward the orientation of the spatial energy. These errors will occur only at angles adjacent to the energy loaded orientation. When stimuli with uniform two-dimensional fourier spectra are moved, no such systematic error should be observed.

METHOD

Displays were created using the Apple II computer driving a Sony Model VM4206 raster scan monitor with P4 phosphor. The monitor face was viewed through a circular mask. The mask and display face were illuminated by a 20 watt incandescent bulb and were of approximately the same brightness and relatively free of shadows. The illumination and low contrast-brightness settings helped eliminate phosphor persistence acting as a cue. Also, the arrangement minimizes edge effects on perception of the stimuli (Kelly, 1977). Stimulus contrast was approximately 10 times visual threshold as estimated using neutral density filters. Viewing distance was from 57 cm so that the visible screen subtended a diameter of 14 degrees of visual angle. A central fixation point was created by the computer and was always visible to help maintain fixation.

Stimuli consisted of simple shapes held in shape tables and moved by successive DRAWing and XDRAWing commands from APPLESOFT BASIC. Before each trial, the computer determined a path which would take the shape through the central fixation point with half of its trajectory either side of that point. The travel covered an average of 5.6 degrees of the central screen area. As mentioned before, to maintain directional control velocity control was sacrificed. Thus, velocities varied between 4.8 and 6.4 degrees per second.

Observers were instructed to maintain their stare at the fixation point while the moving target traversed their visual field. This is difficult to do, so two practice sessions were held to develop this skill. The procedure prevents the observers from using the orientation of the trajectory as a basis for their directional judgement.

Immediately after the stimulus completed its trajectory, a flashing short line segment appeared on the edge of an invisible circle whose diameter was 11.2 degrees of visual angle. Stimulus trajectory never crossed this circle and was limited to a deviation of three degrees from the fixation point. The pointer's initial position was random plus or minus 30 degrees from the actual direction of motion. The observer used the Apple II paddle potentiometer to rotate the pointer about the fixation point to indicate the radius of the imaginary circle which corresponded to the direction of stimulus travel. Resolution of the indicated direction was one degree. If the potentiometer could not be rotated sufficiently, a short press of the paddle button reversed the mapping allowing the observer to report all possible directions. A long button press registered the perceived direction and after a short, variable time interval, started the next trial for self-paced stimulus presentations.

Four independent variables were employed in the experiment. The first variable was directions of motion, of which there were 13 between 0 (rightward) and 180 degrees

(leftward) in approximately 15 degree increments. The actual directions varied slightly due to velocity and direction constraints of the Apple II's relatively low resolution HIRES graphics (279 x 192 resolution).

Second variable was the three shapes were used as moving targets. One was a triangle which was as near to isosceles as the computer screen permitted. The second was an octagon, which was as close to a circle as was possible. A circle would have a two-dimensional fourier spectra with equal energy in all directions. The last shape was a pattern of random dots which subtended a 1.5 degree square area, about the same size as the other stimuli. Dot position within the area was determined by a random number table. Random dots also have a two-dimensional fourier spectra with equal energy in all directions. These last two shapes served as control for possible confounds.

The third independent variable was shape rotation. Shapes were presented either in their originally defined form or rotated approximately 30 degrees clockwise. Some smearing and modification of the shapes occurred because of the rotation process and the Apple II's manner of accomplishing it, but the shapes maintained their basic form.

The last independent variable were the subjects, since a repeated measure procedure was employed. There were three different subjects in the experiment, resulting in three levels of this random variable.

Each observation session consisted of 52 trials. Four judgements were made at each of the 13 directions within each session. Only one shape and rotation were employed during a session. Three sessions were run at each of three shape and two rotation combinations, requiring a total of eighteen sessions. Order of shape-rotation combinations was random and different for each observer. This resulted in 12 observations under each of the independent variable combinations.

RESULTS

There was a great deal of variation in the pattern of perceived direction between the different observers. The results in one of the conditions where systematic perceptual errors were expected, the rotated triangle condition, are plotted in Figure Two.

Insert Figure Two Here

The experienced and non-naive observer, WPM, exhibited the anticipated perceptual error. At directions less than the orientation of the triangle vertices, he overestimated the perceived direction. At directions greater than the orientation of the vertices, he underestimated perceived direction. The other two naive, inexperienced observers did not exhibit a similar pattern of error, contrary to the vector model expectations.

Because of the inconsistency of the signed error data, further analysis used the absolute error to see if more error occurred with the triangle than the other shapes. The mean absolute error was analyzed using a $3 \times 2 \times 13 \times 3$ analysis of variance, repeated measure design. The resulting summary table is presented in Table A.

Insert Table A Here

Unfortunately, the predicted shape effect did not achieve significance. In fact, the only effect to achieve significance was a direction effect ($p < .01$). In order to understand this direction effect, the absolute errors of all three subjects were averaged and plotted in Figure Three.

Insert Figure Three Here

Most accurate perception of perceived direction occurred at the horizontal and vertical orientations. The poorest accuracy seemed occur at the off axis diagonal directions.

Directional accuracy is clearly anisotropic, unequal in all directions. This contradicts earlier observations by Marshak (1981) with full field random dot patterns that accuracy was the same in all directions. The effect was most pronounced in the two inexperienced cadet observers. It may be their inexperience, or in the difference of the stimuli that resulted in the pronounced anisotropy.

DISCUSSION

The results were not at all supportive of the vector model. There are several possible reasons why the predicted outcome did not occur. First, it is doubtful that a single shape had a significant effect on the two-dimensional fourier spectrum of the target. If this were true, systematic errors would not occur. Unfortunately, image analysis was not possible to support or deny this possibility. The appropriate image analysis must be performed on the stimuli to determine if this is true. Also, the experiment must be replicated with stimuli which more radically alter the distribution of spatial energy and at a range of velocities to better capture the pattern-motion interaction effects.

It may also be that the unanticipated directional accuracy anisotropy obscured the effect. Certainly, following experiments will further examine the directional accuracy anisotropy. If its existence is confirmed, then the anisotropy must be taken into account when choosing orientations and directions of moving targets.

In conclusion, the vector model proposed in the introduction is not supported by the data from the present experiment. However, that model's success in explaining and predicting other perceptual phenomena (Marshak, 1981) indicates further, more refined tests of the motion-pattern prediction should be made. The convergent theory and experimental data in the literature and my prior work all suggest that perceived direction of motion can be influenced by a moving object's shape.

Bibliography

Kelly, D.H. Visual contrast sensitivity, *Optica Acta*, 1977, 24, 107-129.

Lettvin, J., Maturana, H., Pitts, W. and McCulloch, W. Two remarks on the visual system of the frog., in W.A. Rosenblith (Ed.), *Sensory Communications*, Cambridge: M.I.T. Press, 1961.

Marshak, W. Perceptual integration and differentiation of directions in moving patterns. Dissertation, Northwestern University, 1981.

Marshak, W. and Sekuler, R. Mutual repulsion between moving visual targets, *Science*, 1979, 205, 1399-1401.

Mather, G. The movement aftereffect and a distribution-shift model for coding the direction of visual movement, *Perception*, 1980, 9, 379-392.

Movshon, J.A. Directionally selectivity in cortical complex cells, Presented at the Association for Research in Vision and Ophthalmology meetings, Orlando, Florida, 1980.

Pantle, A. Motion aftereffect magnitude as a measure of the spatio-temporal response properties of direction-sensitive analyzers, *Vision Research*, 1974, 14, 1229-1236.

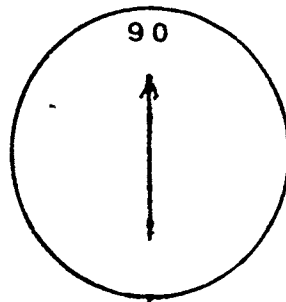
Sekuler, R. and Levinson, E. The perception of moving targets, *Scientific American*, 1977, 236, 60-73.

Ullman, S. *The Interpretation of Visual Motion*, M.I.T. Press, Cambridge, Massachusettes, 1979.

Figure 1

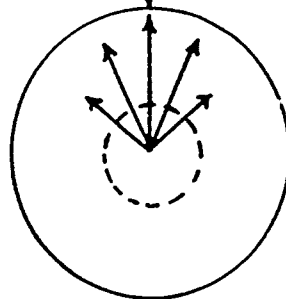
- A schematic representation of the vector model for the perception of motion.
- A. Motion is portrayed in a polar coordinate system. A stimulus moving through 90 degrees is presented to the model.
- B. The base line or unstimulated activity level of the directionally sensitive units is represented by the dashed circle. Only representative vectors exceeding this activity level are represented. Each unit sensitive to 90 degree motion increases its activity in proportion to where that direction intersects the unit's tuning function. Unit activity is represented by a vector lying in the unit's preferred direction and whose length is proportionate to activation.
- C. At the same or subsequent level, active units exert inhibition on other units whose preferred directions are similar. This results in reduced activity in most units, but more pronounced in units whose direction is different from the direction of motion.
- D. The resulting distribution of activity is the basis for perceived direction. Two alternative decision mechanisms are proposed. Either the longest vector (peak) or average of the vectors (vector sum) determines perceived direction.

STIMULUS DIRECTION



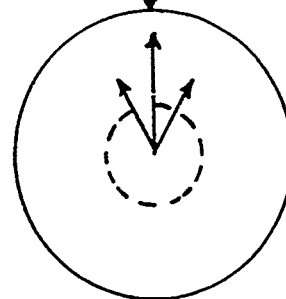
A

EXCITATION



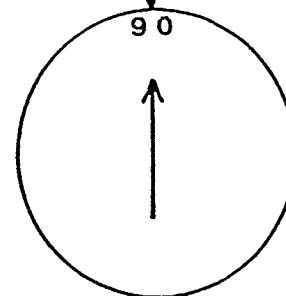
B

INHIBITION



C

DECISION



D

PERCEIVED DIRECTION

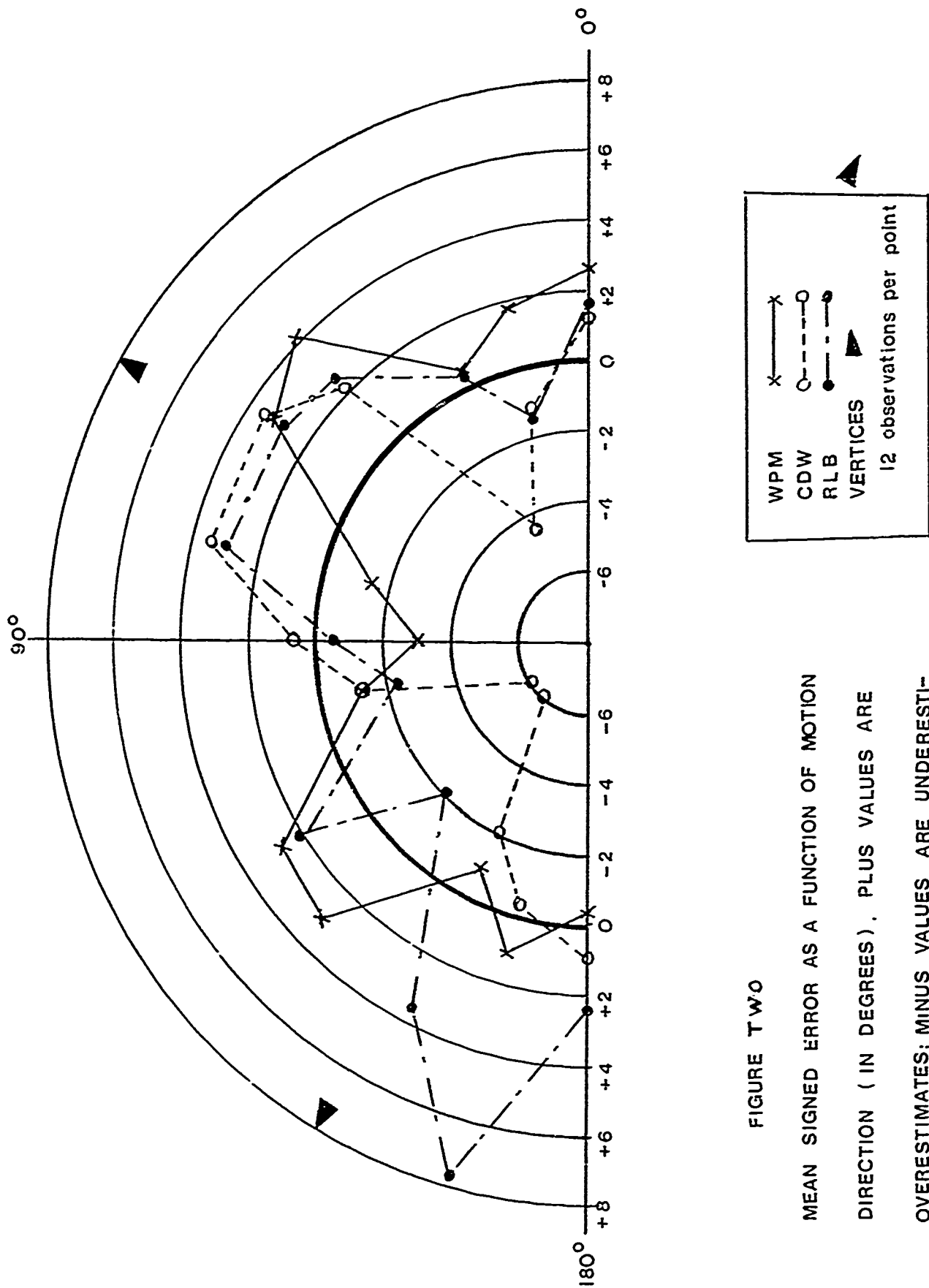


FIGURE TWO
 MEAN SIGNED ERROR AS A FUNCTION OF MOTION
 DIRECTION (IN DEGREES), PLUS VALUES ARE
 OVERESTIMATES; MINUS VALUES ARE UNDERESTI-
 MATES. STIMULUS WAS ROTATED TRIANGLE.

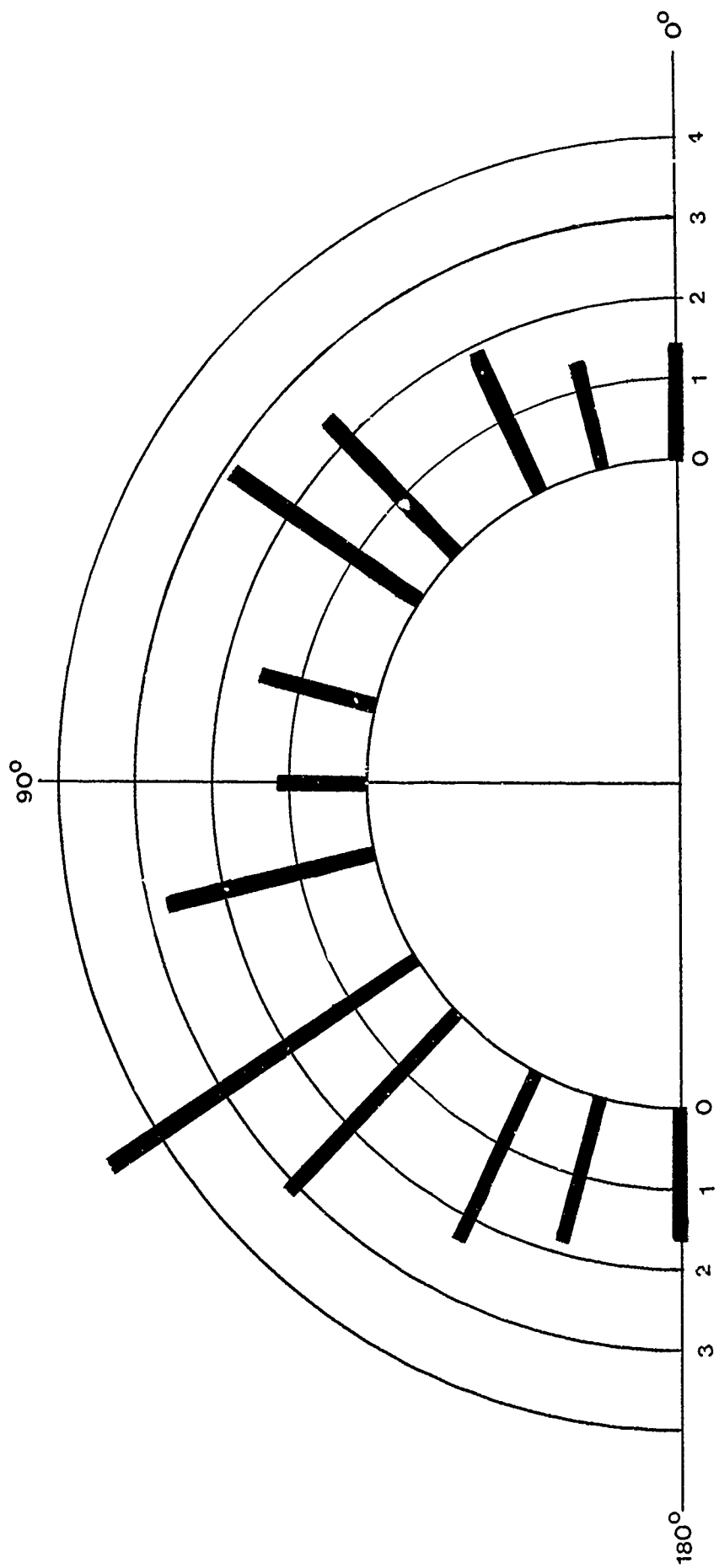


FIGURE THREE
 MEAN ABSOLUTE ERROR AS A FUNCTION OF MOTION DIRECTION (IN DEGREES)
 36 observations per point

SOURCE	DF	SS	MS	F	P
Target	2	6.125	3.063	2.455	n.s.
Subject x Target	4	4.660	1.247		
Target x Rotation	2	1.370	.685	.287	n.s.
Subject x Target x Rotation	4	9.54	2.385		
Target x Rotation	24	49.92	2.080	.811	n.s.
Subject x Target x Direction	48	123.090	2.564		
Target x Rotation x Direction	24	76.770	3.198	1.720	n.s.
Subject x Target x Rotation x Direction	48	88.920	1.850		
Rotation	1	.90	.90	.24	s.
Subject x Rotation	2	3.920	1.960		
Rotation x Direction	12	38.020	3.168	1.570	n.s.
Subject x Rotation x Direction	24	48.600	2.008		
Direction	12	196.93	16.410	3.443	p < .01
Subject x Direction	24	114.400	4.767		
Subject	2	3.140	1.570		

Table A

Analysis of variance summary table for mean absolute
perceptual error.

Proposed New Vision Standards: Contrast Sensitivity

David W. Evans and Arthur P. Ginsburg
Aviation Vision Laboratory
Human Engineering Division
Air Force Aerospace Medical Research Laboratory
Wright-Patterson Air Force Base, Ohio

The Air Force mission is to fly and fight and today the most important component of this mission is the human operator. While many human senses are used during flight, the visual system takes major responsibility for information processing. Even with the advancement of radar and electro-optical sensing devices, visual acquisition is still the most important determinant for success in air-to-air combat. He who detects the enemy first has the best chance of survival and success in a combat situation. Furthermore, superior visual performance can reduce workload by reducing the amount of visual search needed to perform a particular task. Consequently, it is important that pilots, as well as other personnel involved in visual task performance, have the best possible visual capability. To ensure this, stringent visual standards should exist.

Present visual standards rely primarily on the acuity test. Acuity measures visual resolution of small high-contrast letters or objects. This provides a single number for visual capability (e.g., 20/20 or 6/6). Although the acuity measure is quite useful for certain situations, when compared to other types of sensory testing, such as auditory testing, its limitations become apparent. Whereas an audiogram provides sensitivity over the wide range of temporal stimuli normally encountered in everyday situations, visual acuity provides information on sensitivity to only one range of spatial stimuli, small high-contrast objects. An audiologist would scoff at a hearing test that measures sensitivity to only one high-intensity, high-frequency tone, yet a single number has been used for many years as the unitary measure of visual capability. It is no wonder that researchers (Sivak et. al. 1981) have found little relationship between visual performance based on acuity and actual visual task performance. What is needed, if relevant visual standards are to exist, is a measure that relates directly to visual task performance. This measure must test across the wide range of object sizes and contrasts encountered in the everyday visual world. Here, contrast sensitivity is proposed as a useful measure of visual capability.

Contrast sensitivity tests vision in terms of the fundamental spatial frequency component; the spatial sine wave. Varying spatial frequency sine-wave gratings are displayed on a screen and the subject adjusts the contrast until each grating is just detectable. A typical contrast sensitivity function is shown in Fig. 1. The x-axis is the spatial frequency of the sine-wave grating in units of cycles per degree of visual angle, while the y-axis is contrast sensitivity or the reciprocal of the contrast just needed to detect the grating. Contrast sensitivity, then, is a measure of the visibility of different size objects under varying contrast conditions.

Figure 1 about here

Ginsburg (1977) demonstrated the power of the contrast sensitivity technique to detect gross deficiencies in visibility where standard measures could not. For example, Figure 2 shows contrast sensitivity of a multiple sclerosis patient, C.H., who complained of the visibility in one eye as compared to the other. The optometrist found both eyes to be the same when measured with Snellen acuity, 6/6 (European equivalent of 20/20) yet contrast sensitivity showed a factor of four difference between eyes at certain spatial frequencies. This reduction in visibility, undetected by Snellen acuity, was quantified by contrast sensitivity.

Figure 2 about here

If such large differences in visibility go undetected by standard measures, one may question to what extent smaller deficiencies also go undetected. To determine this, individual differences within the population must be quantified.

Individual Differences in Contrast Sensitivity

Figure 3 shows the results of one of the first studies testing individual differences in contrast sensitivity. The contrast sensitivity

Figure 3 about here

functions of three pilots are shown (Ginsburg, 1978). Although all three pilots are similar in Snellen acuity, ranging from 20/15 to 20/25, there are significant differences in contrast sensitivity below 7 cycles/degree. Furthermore, the pilot who had the worst Snellen acuity, 20/25, was a factor of three more sensitive than the 20/20 subject at the peak spatial frequency, four cycles/degree. Similar results have been found with large numbers of subjects. Figure 4 (Ginsburg et. al., 1980) shows data collected on 265 subjects. The thick black curve is the median while the dashed and thinner black curves capture the 75th and 90th percentiles, respectively. A factor of three difference between the 90th percentile curves exists at the peak spatial frequency. The difference at the lower and higher spatial frequencies are even greater. This result shows that

Figure 4 about here

large individual differences in contrast sensitivity exist across the population. To what extent these individual differences in contrast sensitivity affect visual task performance is yet another question.

Predicting Visual Performance

For any measure to be relevant to the operational community, it must be able to predict performance. In 1981, Ginsburg and Evans showed contrast sensitivity could be used to predict target acquisition performance of static targets. The contrast sensitivities of seven Air Force personnel were measured along with their detection and identification performance for Snellen letters and aircraft silhouettes. Correlations between contrast sensitivity and performance were quite high ($\geq .73$). Standing alone, these correlations are an important result, yet they do not typify the true predictive power of the contrast sensitivity measure. In the correlations, all subjects were grouped together and individual differences were de-emphasized. A striking example of individual difference appeared between two subjects in this study as contrast sensitivity and Snellen acuity predicted opposite results. Subject A3 has 20/20 binocular Snellen acuity while A5 has 20/30 binocular Snellen acuity (subjects were tested without corrective lenses). If target acquisition performance could be determined by Snellen acuity, then A3 would have been predicted to do better than A5. This is not the case. Comparing contrast sensitivity functions (Fig. 5) of A3 and A5, we find that A5 is much more sensitive in the mid-frequency region of the curve (two to twelve cycles/degree). Accordingly, A5's performance for detection and iden-

Figure 5 about here

tification of the large and medium size letters and aircraft silhouettes is much better than subject A3. , performance on the smaller, higher contrast letters and aircraft silhouettes, A3 has better performance. This is expected, since their contrast sensitivity curves cross in the high frequency region (smaller objects) and A3 becomes more sensitive than A5. This demonstrates that in order to predict visual performance, a test, such as contrast sensitivity, which measures across a wide range of object sizes and contrasts, must be used.

The previous study was limited to static conditions. A more demanding test requires prediction of visual performance in a complex real-world task. In 1981 Ginsburg, Evans, Sekuler and Hays ran the first study relating individual pilot contrast sensitivity to air-to-ground target detection in a visual flight simulator.

The photopic (210 cd/m^2) and scotopic ($.21 \text{ cd/m}^2$) Snellen acuities and photopic (150 cd/m^2) and scotopic ($.15 \text{ cd/m}^2$) contrast sensitivities of 11 Air Force instructor pilots were tested. They also performed an air-to-ground detection task in a simulated A-10 instrument landing approach. The pilots were not flying the simulator but acting as passive observers, as the visual scene was slaved from another simulator being flown by another pilot. The task was to detect a MIG aircraft positioned at the end of the runway. The MIG was presented at random so that the pilot did not know in which trials the MIG would be present.

Figure 6 about here

The highest correlations were found between photopic and scotopic peak contrast sensitivity and air-to-ground detection performance; $r=.24$ and $r=.83$, respectively. The lowest correlations occurred between photopic and scotopic Snellen acuity and the detection task, $r=-.24$ and $r=-.13$, respectively. The higher correlation found with the low luminance contrast sensitivity was expected since the luminance in the simulator (1.9 cd/m^2) more closely matched the scotopic contrast sensitivity luminance. Figure 6 shows both photopic and scotopic contrast sensitivities of the pilots and the scatter diagrams for detection performance and scotopic contrast sensitivity and Snellen acuity. Even though both contrast sensitivity and Snellen acuity are matched for luminance, contrast sensitivity predicts performance; Snellen acuity does not. This result again demonstrates predicting visual performance requires measuring vision across varying object sizes and contrasts.

Conclusions

This brief paper has presented data to show that contrast sensitivity can be used as a relevant measure of visual capability. Not only can it detect individual differences in visual capability, but it can relate these differences to task performance. We suggest that further contrast sensitivity data be collected across DoD personnel in conjunction with the conventional acuity measure. These data can then be used to develop standards that relate directly to task performance. In turn, these standards will help optimize the most important aspect of the Air Force mission: the human operator.

References

Ginsburg, A.P. "Visual Information Processing Based on Spatial Filters Constrained by Biological Data," Ph.D. Dissertation, University of Cambridge, 1977.

Ginsburg, A.P. "Spatial Filtering and Vision: Implications for Normal and Abnormal Vision," in Applications of Psychophysics to Clinical Problems, (Proceedings of the Symposium held in San Francisco, CA, October 1978, in press, The Cambridge University Press).

Ginsburg, A.P. and Evans, D.W. "Relationships between Individual Differences in Contrast Sensitivity and the Detection and Identification of Targets," (Proceedings of the Aerospace Medical Association Meeting, held in San Antonio, TX, May 1981).

Ginsburg, A.P., Cannon, M.W., Sekuler, R., Evans, D., Owsley, C. and Mulvanny, P. "Large-Sample Norms for Contrast Sensitivity," in press, Investigative Ophthalmology and Visual Science, 1981.

Ginsburg, A.P., Evans, D.W., Sekuler, R. and Harp, S. "Contrast Sensitivity Predicts Pilots' Performance in Aircraft Simulators," in press, American Journal of Optometry and Physiological Optics, 1981.

Sivak, M., Olson, P. and Pastalan, L. "Effect of Driver's Age on Nighttime Visibility of Highway Signs," Human Factors, 23(1), 59-64, 1981.

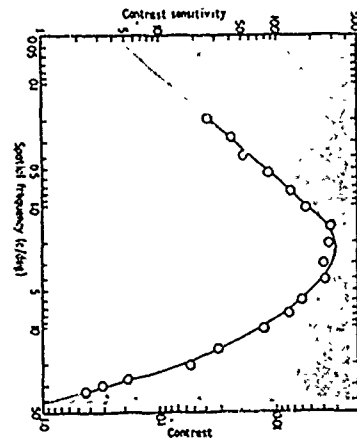


FIG. 1
Contrast sensitivity function (from Ginsburg 1977).

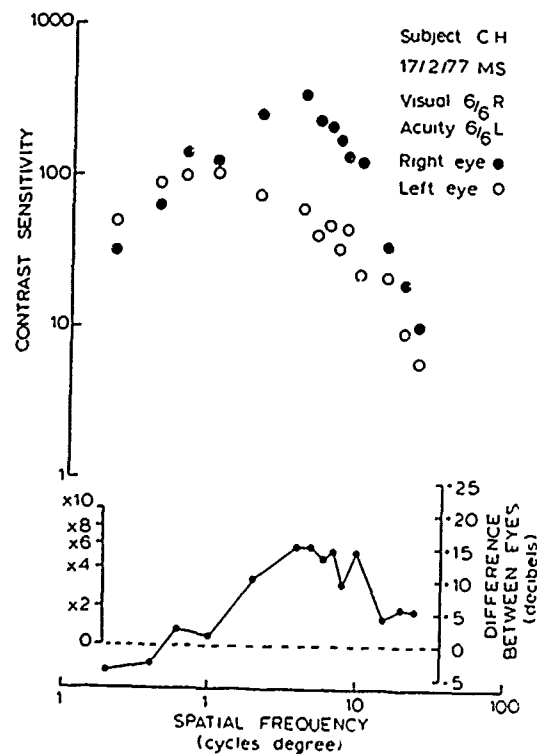


FIG. 2
Contrast sensitivity function of subject with multiple sclerosis. Lower difference curve shows factor of four difference between eyes over a log unit range of spatial frequency (from Ginsburg 1977).

INDIVIDUAL DIFFERENCES IN DETECTION
SENSITIVITY FOR SINE-WAVE GRATINGS
OF THREE AIR FORCE PILOTS

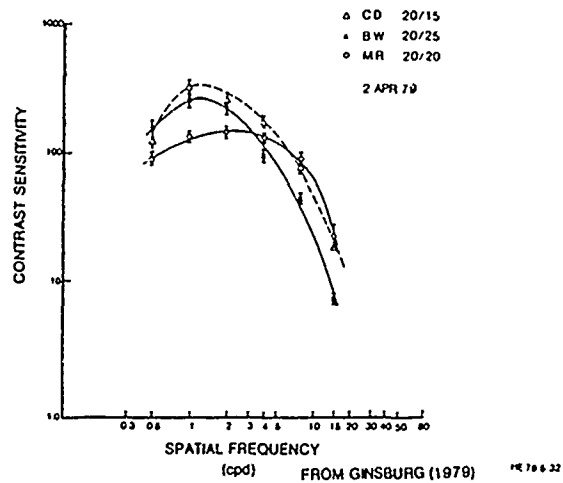


FIG. 3

The contrast sensitivity functions of three pilots with Snellen acuities of 20/15, 20/20 and 20/25. (from Ginsburg 1977).

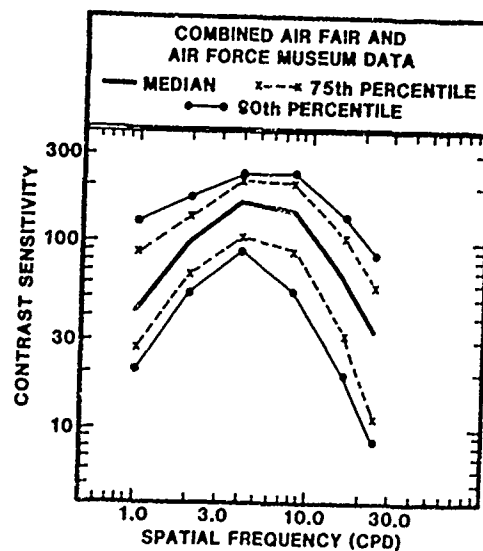


FIG. 4

Median, 75th percentile and 90th percentile contrast sensitivity curves for data on 265 subjects (from Ginsburg et. al., in press 1981).

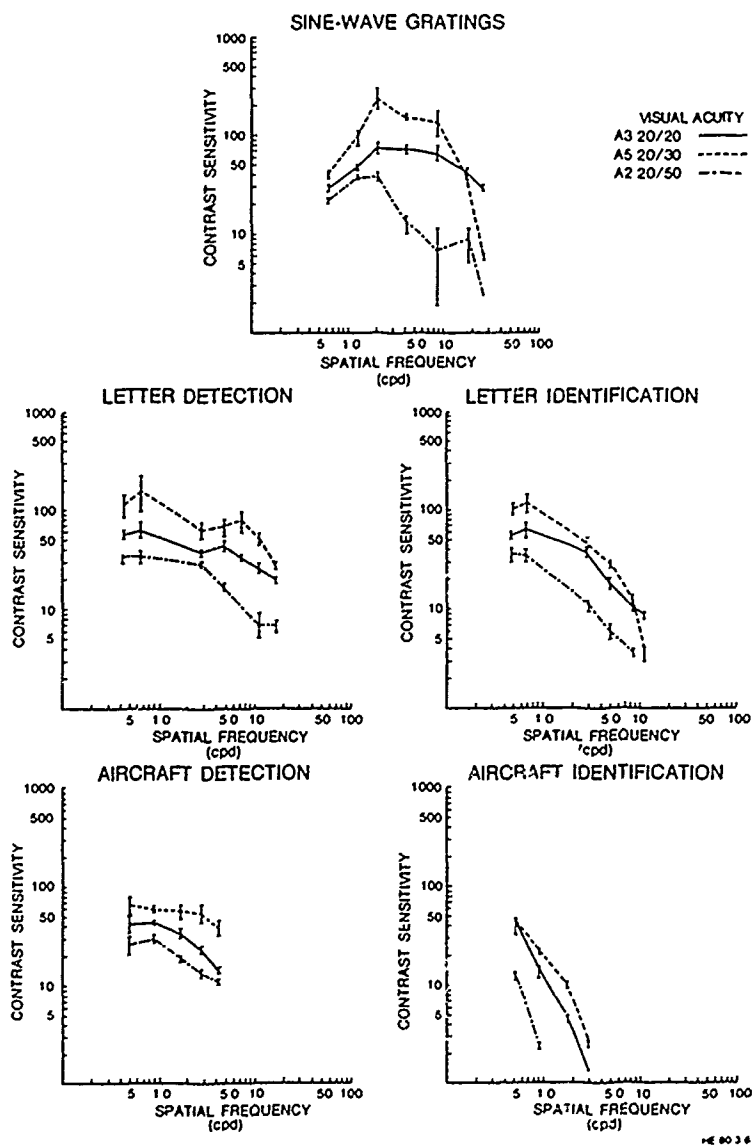


FIG. 5
Contrast sensitivity to sine-wave gratings, and detecting and identifying letters and aircraft silhouettes for subjects having 20/20, 20/30 and 20/50 binocular Snellen acuity. (from Ginsburg and Evans 1981).

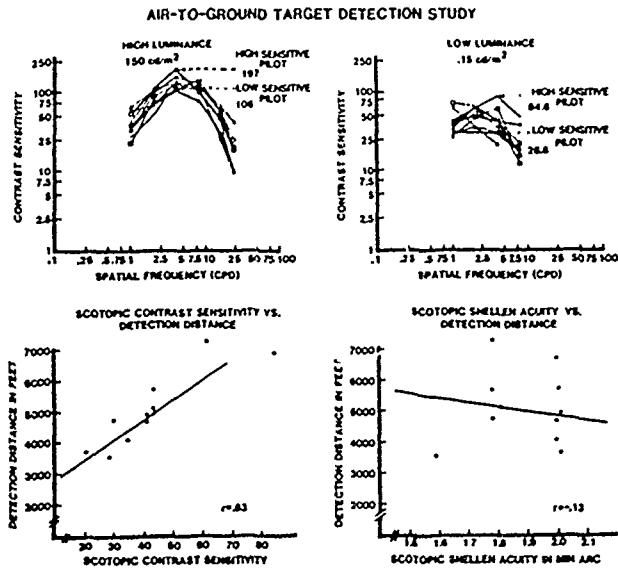


FIG. 6

The uppermost curves show the contrast sensitivity functions of 11 Air Force pilots taken at high and low luminance. The most and least sensitive pilots are shown in heavy black lines. The correlation between the contrast sensitivity ($r = .83$, $p < .01$) and Snellen acuity ($r = -.13$) of the pilots and their detection range for a ground target, lower plots, shows contrast sensitivity predicts performance, visual acuity does not. (from Ginsburg, Evans, Sekuler and Harp, in press 1981).

PILOT SELECTION FOR TARGET DETECTION

Russell A. Benel and Denise C.R. Benel

Essex Corporation

Abstract

The tacit assumption that the observer with "normal" vision has a relaxed state of accommodation at optical infinity has been shown to be false in most specific instances. Evidence suggests that, in the absence of patterned stimuli, observers assume a characteristic intermediate position of visual accommodation (commonly referred to as the dark focus). Since individuals vary in the precise position of their dark focus and the dark focus appears predictive of certain aspects of visual performance, it is likely that use of the measured dark focus might improve the selection and classification of pilots. This paper reviews the available data and presents a program to determine the potential use of the dark focus for pilot selection and classification.

A puzzling and persistent problem has been the manifestation of inappropriate visual accommodation, i.e., the observer is focused nearer or farther than would be expected. Night myopia has been known for at least two centuries, with the report in 1789, by Lord Maskelyne, the royal astronomer, that the use of a negative lens facilitated his night observation. "Empty field myopia" (or "space myopia") has been discussed in detail by Whiteside (1957) with particular reference to high altitude flight. Also, when looking through microscopes, observers typically exhibit unnecessary increases in accommodation referred to as "instrument myopia" (Hennessy, 1975).

The heuristic explanation can be described by the intermediate resting state of accommodation after Schober's (1954) Akkommodationsruhelage. This is merely a descriptive statement that the eye has a natural tendency to focus at an intermediate position approximately at arm's length. The alternative, traditional view (after Helmholtz, 1867/1962) is that the resting position of accommodation is at optical infinity. This theory was consistent with several assumed properties of the visual system. Aside from the scientific arguments, there is a semantic impetus for the optical infinity or zero diopter (diopters are the reciprocal of the focal distance in meters) resting position. As Morgan (1957) cogently noted, the absence of activity is best represented by zero. Evidence to support an intermediate resting distance is now ample.

Anomalous Myopias

The anomalous myopias have a long, but equivocal history. For example, night myopia has been reported many times over the past 200 years, but its cause has been in dispute. Liebowitz and Owens (1975) provided substantive evidence for an accommodation-based explanation finding the measured dark focus (the focus of the eye in absolute darkness) to correlate quite highly ($r = .84$) with the observers visual response to a simulated twilight view of a distant building. They also demonstrated a similar correlation between the dark focus and empty-field myopia. Empty-field (space or sky) myopia is manifested when viewing an unstructured field (Ganzfeld) such as a clear sky. The potential impact of this phenomenon is most clearly exemplified by Whiteside's (1957, p. 67) description of air-to-air search above the clouds:

... the direction which it would appear was known. In spite of this help, when the target aircraft was (ital. his) seen it was almost invariably detected clearly and suddenly and was much nearer than would have been expected ... the impression of the difficulty in focusing was so strong as to give rise to a sensation of disorientation such as is sometimes experienced when one is in total darkness.

In contrast to the previous myopias, instrument myopia occurs during observation of targets of high contrast and rich detail. Hennessy (1975) reported that luminance, magnification, wave length, and visual angle of the field have all been varied with little effect on the magnitude of instrument myopia. He found that the mean refractive state while viewing a square-wave target through the microscope was 1.91 D. This corresponded closely with the observers' responses in the dark ($r = .78$). Hennessy concluded that instrument myopia and the dark focus are manifestations of the same phenomenon, the resting state of accommodation.

The Mandelbaum Effect

Frequently, observation of distant targets occurs through an interposed surface, e.g., a windscreen. Hence, the eye may focus either the interposed surface or the target object. Whiteside (1957) observed that some pilots focused spots on the aircraft canopy during air-to-air search, but Mandelbaum (1960) was apparently first to document the circumstances under which the interposed texture was focused rather than the desired object. He subsequently demonstrated informally that involuntary accommodation to the screen caused acuity loss, also noting large individual differences in the critical subject-to-screen distance. Owens (1979) and Benel (1979) found that the "Mandelbaum effect" is greatest when the interposed screen and the observer's dark focus correspond, i.e., the eye exhibited an inherent bias toward objects presented near the dark focus.

The Dark Focus

A proper test for a "resting" state can only be conducted when the lowest level of ambient stimulation is present. Morgan (1957) proposes two possible conditions in which visual stimuli are reduced to the minimum — in complete darkness and in a luminous but completely empty visual field. The exact value that the eye assumes varies widely among observers. Leibowitz and Owens (1975) collected accommodative responses in the dark from 124 college students and found the mean DF to be approximately 1.7 D (59 cm). All observers had at least 20/25 corrected near and far acuity. Only four observers had a DF of 0.5 D or less. According to the classical view, the majority should have had responses in darkness corresponding to optical infinity. The distribution is approximately normally distributed with a standard deviation of 0.72 D and a range of 4 D. Leibowitz and Owens (1978) replicated the essential findings of their earlier study with 220 college students who had a mean measured DF of 1.5 D. Simonelli (1979) provided evidence for the profound sampling effects that may occur in these measurements. Between college students and US Air Force recruits he demonstrated a difference of 1.5 D in the measured DF (approximately 1.2 D for recruits and 2.7 D for students).

Potential Selection Measures

Marano (1981) reviewed the history of visual acuity testing at the 1981 Meeting of the Aerospace Medical Association (ASMA). A number of "standards" were covered including the ubiquitous Snellen chart from which we have derived our measure of 20/20 vision. This measure is based on "the minimum visible and the minimum separable of an object." A variety of alternatives have been proposed over the years since Donders first presented Snellen's test types in 1862. Nearly all alternatives have been based on the minimal discriminable form at a given distance.

At this same ASMA meeting Ginsburg and Evans (1981) presented a cogent argument for improving visual standards by considering additional visual parameters. To quote, "Visual standards based on an ability to see small, high-contrast letters or symbols are inadequate to evaluate visual capability for target acquisition over ranges of target size and contrast found in operational environments." They propose to incorporate the measure of contrast sensitivity which uses the visibility of sine-wave gratings to assess

visual capability. This measure predicts target detection and identification performance better than standard acuity measures. Owens (1980) compared accommodative responsiveness and contrast sensitivity for sinusoidal gratings noting that steady-state accommodation is a linear function of object distance. Thus, one index of stimulus effectiveness is the slope of the accommodative response function (see Benel, 1980). Two points are evident in Owens' data. One is that contrast resolution and accommodative responsiveness share common spatial response characteristics. This may be seen in the high correlation between accommodative responsiveness and contrast sensitivity and the tendency for accommodation to return to the DF whenever the gratings would appear to fade (due to adaptation).

The DF itself is a relatively new measure in the arena of individual visual characteristics. The DF has been shown to be highly predictive of visual functioning of settings. For present purposes, the most interesting correlate of the dark focus is target detection performance. In Owens' (1980) work sine-wave gratings disappeared when observers accommodation was not accurately matched to the presentation distance. Whiteside (1957) reported the inability of pilots to detect objects in empty fields until they were very near. Post, Owens, Owens and Liebowitz (1979) reported the possibility of improved detection performance in empty fields by "correcting" the DF which is functionally equivalent to selecting those with a more distant DF. Thus, it seems that a distant DF would be a useful attribute for those attempting such tasks as air-to-air search.

Stability of the Dark Focus

An important consideration before utilizing the DF in selection, prediction, etc., is its temporal stability, both "trait" stability and "state" stability. The former is indicative of a long term invariance in the measure — something characteristic of the person. The latter is indicative of temporary fluctuations about such a trait level.

Evidence of DF shifts with mood is limited. Miller (1978) actually measured DF, but the state shifts showed trends only — no strong relationships. Similarly, Westheimer (1957) (investigating the resting position in empty field) angered subjects by insults from the experimenter and observed a rise in accommodation lasting several minutes. This does not refute the DF is a reliable measure of visual functioning. In fact, similar results are often found in the time-honored acuity measures. In terms of the potential for pilot selection/classification a cautionary note would be in order. Just as those with marginal acuity might fail to meet acuity standards, it would be expected that a number of individuals with a marginal DF might fail to prove acceptable on one measurement instance, but would "pass" on a second. Whether individuals with this type of lability should be acceptable or not is an empirical question.

As for trait stability, Miller (1978) reported stability over two to three weeks. His subjects, however, were measured only two days each week, morning and afternoon, yielding a maximum of twelve measurement times per subject. Owens and Higgins (1979) reported individual differences in the magnitude and pattern of variability of the DF, but the subjects oscillated about a stable mean DF over a one-year period. Mershon and Amerson (1980) have shown that the variability of the DF can be less than ± 3 D when time of day is controlled.

Measuring the Dark Focus

Numerous methods and instruments have been proposed for measuring the refractive state of the eye. In clinical practice retinoscopy is one common method and has been adapted to the measurement of the dark focus. The instrument requires considerable training and practice to use. A number of alternative devices have been developed to measure accommodation objectively. Some of these devices are quite complex and

expensive (e.g., the SRI infrared optometer and eye tracker). Although the more complex devices may provide continuous readout of accommodation, for dark focus measurement a momentary reading is sufficient. A second drawback to continuous recording optometers is the difficulty of calibrating an observer's response to optical infinity.

The laser optometer (Leibowitz and Hennessy, 1975) overcomes this latter drawback. By using the Badal principle a compact, portable device may be constructed which provides an objective measure of the true refractive state. In practice, a bracketing procedure is used to determine the distance of accommodation. One major advantage of this device is the fact that the laser speckle pattern does not serve as an accommodative cue. Therefore, unconfounded measures of the dark focus are possible. The device requires modest training for users, is relatively unobtrusive, and has an historical precedent for its use.

If the laser optometer has any drawbacks at all, they would be the use of a laser and the possibility that some individuals, for unknown reasons, can't be measured by this technique. Simonelli (1979) describes an alternative device that is elegant in its simplicity. It incorporates the Scheiner principle through use of split polarizing filter material as a vernier. Rather than report apparent speckle flow as in laser optometry, the observer merely reports alignment or misalignment of the vernier bars. Simonelli's device incorporated a shutter (to reduce exposure time so as not to allow accommodation to change) and the Badal principle to allow a compact size. Also, the device is quite accurate with resolution approaching 0.1 D. In a comparison with the laser optometer, Simonelli found a high correlation between the two devices ($r = .93$) and no reliable difference between readings. He also determined a high test-retest reliability for his device ($r = .94$). In subjective ratings of confidence of response and ease of use the vernier optometer was favored. Although there may be a number of circumstances for which the laser optometer would still be the instrument of choice, for simple dark focus measures the polarized vernier optometer holds many advantages.

A Research Program Plan

First, a clinically reliable screening instrument should be fully developed. Simonelli's device appears most promising from the standpoint of cost and ease of use. Several changes should be incorporated in the device including the ability to measure each eye individually and a rapid method of aligning the pupil with the split polarized material. These changes would allow a larger number of people to be measured in given period. A relatively large normative sample should be tested. The current evidence is insufficient to provide reasonable bounds for the actual distribution of this visual characteristic in the population. Also, there is evidence for a change in distribution characteristics with the mean age of the cohort being measured. Additional data should be collected on test-retest reliability. The reliability should be no worse than that of other visual standards (i.e., acuity tests). There may be individual differences in reliability with implications for target detection performance. Research must be conducted to elucidate the relationship between detection performance and the DF. Thus, some estimate of detection performance confidence limits would be established. The evidence suggests that detection performance would be enhanced for those with a distant DF. The relationship between the improvement due to the DF and that reportedly related to contrast sensitivity is unknown. Therefore, additional work is required on the relation between the DF and other measures of visual functioning (e.g., contrast sensitivity, acuity, and refractive indices). Even with the availability of highly sophisticated electronic detection and tracking equipment, the pilot with initial visual acquisition of his target will be likely to survive in air-to-air combat. The competitive edge will go to the pilot with first sight capability. Whether this comes through improved selection, visual training, and/or prosthetic devices (glasses) is immaterial. The absence of that capability is an unacceptable risk. A risk that would be eliminated through this research.

REFERENCES

- Benel, R. A. Visual accommodation, the Mandelbaum effect, and apparent size. Ph.D. dissertation, University of Illinois at Urbana-Champaign, October 1975. (Short version in Proceedings of the 24th Annual Meeting of the Human Factors Society, Los Angeles, CA: October 1980.)
- Benel, R. A. A functional metric for the description of visual stimuli. Proceedings of the 7th Psychology in the DOD Symposium. USAF Academy: April 16-18, 1980.
- Ginsburg, A. P. and Evans, D. W. Toward meaningful visual standards: Relating individual visual capabilities to target acquisition. Preprints of the Annual Meeting of the Aerospace Medical Association, San Antonio, TX: May 4-7, 1981.
- Helmholtz, H. Physiological Optics. New York: Dover, 1867/1962.
- Hennessy, R. Instrument myopia. Journal of the Optical Society of America, 1975, 65, 1114-1120.
- Leibowitz, H. W. and Hennessy, R. T. The laser optometer and some implications for behavioral research. American Psychologist, 1975, 30, 349-352.
- Leibowitz, H. W. and Owens, D. A. Anomalous myopias and the intermediate dark focus of accommodation. Science, 1975, 189, 646-649.
- Leibowitz, H. W. and Owens, D. A. New evidence for the intermediate position of relaxed accommodation. Documenta Ophthalmologica, 1978, 46, 133-147.
- Mandelbaum, J. An accommodation phenomenon. Archives of Ophthalmology, 1960, 63, 923-926.
- Marano, J. A. Historical perspective on visual acuity tests. Preprints of the Annual Meeting of the Aerospace Medical Association, San Antonio, TX: May 4-7, 1981.
- Mershon, D. and Amerson, T. L., Jr. Stability of measures of the dark focus of accommodation. Investigative Ophthalmology and Visual Science, 1980, 19, 217-221.
- Miller, R. J. Mood changes and the dark focus of accommodation. Perception and Psychophysics, 1978, 24, 437-443.
- Morgan, Jr., M. W. The resting state of accommodation. American Journal of Optometry and Archives of the American Academy of Optometry, 1957, 34, 347-353.
- Owens, D. A. The Mandelbaum effect: evidence for an accommodative bias toward intermediate viewing distances. Journal of the Optical Society of America, 1979, 69, 646-652.
- Owens, D. A. A comparison of accommodative responsiveness and contrast sensitivity for sinusoidal gratings. Vision Research, 1980, 20, 159-167.
- Owens, R. L. and Higgins, K. E. The long-term stability of the dark focus of accommodation. American Journal of Optometry and Physiological Optics, 1979, 56, 781 (abstract).
- Post, R. B., Owens, R. L., Owens, D. A., and Leibowitz, H. W. Correction of empty-field myopia on the basis of the dark focus of accommodation. Journal of the Optical Society of America, 1979, 69, 89-94.
- Shober, H. Über die Akkommodationsruhelage. Optik, 1954, 11, 282-290. Translation "On the resting position of accommodation" by R. Koeck and H. Leibowitz.
- Simonelli, N. M. The dark focus of visual accommodation: Its existence, its measurement, its effects (Tech. Rep. BEL-79-3/AFOSR-79-7). Las Cruces, NM: New Mexico State University, Behavioral Engineering Laboratory, 1979.
- Westheimer, G. Accommodation measurements in empty visual fields. Journal of the Optical Society of America, 1957, 47, 714-718.
- Whiteside, T. C. D. The problems of vision in flight at high altitude (AGARDograph No. 13). London: Butterworths Scientific Publications, 1957.

EFFECTS OF TASK PERFORMANCE ON VISUAL ACCOMMODATION AND PERCEIVED SIZE

Valerie J. Gawron
Arvin/Calspan
Buffalo, NY

Kenneth R. Paap
New Mexico State University
Las Cruces, NM

ABSTRACT

Previous research has suggested that cognitive tasks tend to induce outward shifts in accommodation that, in turn, result in changes in size perception. To test this hypothesis, we asked 12 people to rest or to perform a running-memory task with auditory stimuli or visual stimuli and to make 4 size judgments under each condition. Comparisons of mean-accommodation measures (obtained using an infrared optometer) and pre- and post-experiment dark-focus measures yielded no reliable differences.

INTRODUCTION

In most systems, operator performance depends on visual information; e.g., an airplane pilot often makes decisions based on what he or she sees beyond the windscreen. Do tasks involving high levels of risk, uncertainty, or difficulty induce internal states (e.g., stress, arousal, high demands on cognitive-processing capacity) that trigger physiological changes directly affecting visual perception?

As shown in the summary of previous research (Table 1), compelling, indirect evidence exists to support the hypothesis that cognitive-task performance tends to induce outward shifts in accommodation that affect the veridicality of an operator's percepts.

RESEARCH STRATEGY AND METHODOLOGY

While the findings of these investigations generally support the hypothesis that task performance results in an outward accommodative shift and that accommodation and perceived size are in some way related, they are open to alternative interpretations. To clarify the roles of other potentially influencing variables, this study imposed 3 controls: (1) to ensure that outward accommodative shifts could not be attributed to either the simple passage of time or the presence of memory-task stimuli, subjects were presented with memory-task stimuli, but were instructed to ignore them; (2) to eliminate the possibility of outward shifts toward the resting position as a result of accommodative fatigue (caused by a subject carefully fixating stimuli and making numerous oculomotor adjustments), the same memory task was presented in both

Table 1
SUMMARY OF PAST RESEARCH

RESEARCH AREA	RESEARCHER(S)	RESULTS	COMMENTS
Relationship between sympathetic nervous system and accommodation	Alpern (1969)	<ul style="list-style-type: none"> Stimulation of sympathetic nervous system results in a flattening of the lens of the eye. 	<ul style="list-style-type: none"> Consistent with Skeffington (1967)
Sympathetic activation, task difficulty and accommodation	Randle, Roscoe, and Pettit (1980)	<ul style="list-style-type: none"> Upon measuring accommodative states of pilots during a simulated night landing, found outward accommodative shift as airplane approached touchdown and flight-decision importance increased. 	<ul style="list-style-type: none"> Supports hypothesis that sympathetic activation produces some flattening of the lens as cognitive demands and/or arousal increases.
Effect of task performance on accommodation	Malstrom, Rendle, and Weber (1975)	<ul style="list-style-type: none"> Found accommodation during dual-task performance (tracking a target presented at 0 to 3D while performing a backward-counting task) more reliably distant than accommodation during single-task performance (tracking only) 	
	Malstrom (1978)	<ul style="list-style-type: none"> Expanded experimental design, i.e., 3 levels of difficulty in mental task (backward counting by 1s, 3s, and 7s) and 2 response modes on secondary task (written or mental), again yielded outward accommodation shift associated with task performance. 	<ul style="list-style-type: none"> Consistent with hypothesis that task performance includes sympathetic activation that stimulates outward shifts in accommodation.
Effects of task performance on sympathetic activation and accommodation	Skeffington (1957)	<ul style="list-style-type: none"> Found greater distance accommodation when subjects read difficult material than when subjects read easy material. Measured elevated blood pressure, respiration rate, and galvanic skin response (GSR) in difficult condition. 	<ul style="list-style-type: none"> Supports hypothesis that amount of sympathetic activation increases with task difficulty. Suggests hypothesis that general sympathetic shifts mediate effects of task performance on accommodation.
	Gawron (1979)	<ul style="list-style-type: none"> Using cognitive processing task similar to Malstrom's and sympathetic-activation indices similar to Skeffington's and measuring multiple physiological and visual responses before and after short-term memory task or rest for equivalent time period, found: (1) task performance associated with sympathetic shift for most variables including diastolic blood pressure and accommodative near point, (2) sympathetic shifts following short-term memory task for skin conductance, heart period, respiration rate, (3) parasympathetic shifts following rest, (4) same interaction pattern approaching statistical reliability for resting position and far point. 	<ul style="list-style-type: none"> Supports hypothesis that a general sympathetic activation is associated with performance of a simple task.
Relationship between accommodation and perceived size	Roscoe, Olzak, and Rendle (1976)	<ul style="list-style-type: none"> Switching from a binocular view of a distant target (2 to 4m) to a monocular view resulted in accommodative shifts inward toward the resting position and a concomitant reduction in apparent size. Switching from a binocular view of a near target (0.25 to 0.50m) to a monocular view resulted in perceived size changes that were either nonsignificant (0.25m target) or unexpectedly reduced (0.50m target). 	<ul style="list-style-type: none"> Suggests other factors are operating at near distances: simple relationship between accommodation changes and perceived sizes should not be expected, particularly for binocular viewing and/or at distances nearer than the resting position.
	Iavecchia, Iavecchia, and Roscoe (1978) Simonelli and Roscoe (1979) Hull, Gill, and Roscoe (1979)	<ul style="list-style-type: none"> Found reliable correlations ($r = 0.89$) between accommodation and monocular judgments of perceived size at distances between 30 and 200m. Found this correlation smaller at night ($r = 0.72$) than during the day ($r = 0.94$). Found no significant differences in either perceived size or accommodation when absolute distance to a real, visible terrain was manipulated while depressed angular position (retinal focus) was held constant. 	<ul style="list-style-type: none"> When stimulus condition produces reasonably large variations in the perceived size of targets at least 1m away, there will be a strong relationship between accommodation and perceived size.

the visual and auditory modes (similar results in both modes favor the cognitive-load/sympathetic-activation explanation over the accommodative fatigue explanation); and (3) to ensure that results would not be confounded by potentiation effects in the extraocular muscles, the visual display was placed at 35 cm [an average Physiological Position of Rest (PPR)] and viewed monocularly. Pre-test and post-test measures were recorded for the perceived straight-ahead view for each eye alone, and accommodation was measured under reduced visual conditions.

Seven men and 5 women, all having normal vision without correction and varying in age from 18 to 30, participated in all four conditions produced by factorially combining the two levels of task variable (i.e., performing or not performing the cognitive task) and the two levels of modality (i.e., aural or visual letter stimuli). Each condition lasted 4 minutes and the order of condition presented was counterbalanced across subjects. Before presentation of each condition, subjects were informed whether they would perform the cognitive task and whether the letter stimuli would be visual or aural. Visual single-letter stimuli were presented for 500 msec at 3-second intervals; aural single-letter stimuli of equal duration and intensity were presented from a high-quality tape through a loudspeaker, also at 3-second intervals. An ADAC LSI-11/03 microprocessor controlled event timing, generated visual stimuli, and recorded subject response. An infrared optometer (Cornsweet and Crane, 1970) was used to measure accommodation, providing continuous, objective, and nonobtrusive estimates of left-eye accommodation.

Cognitive Task. Subjects were told that single letters would be presented visually or aurally and were instructed to make a keyboard response based upon the present letter and the letter one back in the sequence. They were also given a list of words that could be formed by inserting the vowels e, i, or o between successive pairs of stimulus letters. Whenever the preceding and current stimulus letters could be used to form a 3-letter word by mentally inserting e, i, or o, the subject was instructed to press the left key; if no word could be formed, the subject was instructed to press the right key. This is a reasonably difficult cognitive task since the subject must update and hold the last letter presented in short-term memory, mentally insert each vowel into the context formed by the last and current letter, and make a lexical decision about each letter string until a word match occurs or until the three vowels have been exhausted.

Size Judgments and Accommodation Measures. Subjects were instructed to make judgments on a scale from 0 to 100 in relation to a standard. If the line were of equal height to the standard, the judgment would be 100. If it were half as long, 50. Subjects were allowed to see the standard before the experiment and were asked to make 4 size judgments during each condition. Stimulus lines were presented 1, 2, 3, and 4 minutes after the start of each condition. To measure accommodative state at stimulus onset, the line stimulus was marked on a strip-chart recorder, concurrent accommodation and size judgments under dark-focus conditions and accommodation during darkness with no target were also recorded before and after the experiment.

Straight-ahead Tests of Muscle Potentiation. To test for muscle potentiation, monocular judgments of straight-ahead focus were measured using a horizontal string of dots presented on a plasma panel, while subjects were positioned in the biteboard. Subjects were instructed to focus on the dot perceived as being "straight ahead" and to press a key as soon as the focus dot disappeared. Dots were then successively erased at the rate of 4 dots/sec. Based upon ten trials per eye (performed before the 4 experimental conditions), a critical range (1 standard deviation from the mean perceived straight-ahead focus) was calculated for each subject. If the post-test measure fell outside the critical range for that eye, the subject's data was discarded. Consequently, the results of this study reflect little or no muscle potentiation.

RESULTS

Each of 12 subjects (whose pre- and post-test estimates of the PPR for the horizontal-vergence system were within 1 standard deviation) made 4 size judgments in each of the 4 experimental conditions. No significant differences were found.

Using each subject's dark-focus and pre-test accommodation to the line targets as covariates, we performed two analyses of covariance on the mean-accommodation responses. Neither yielded significant effects or interactions.

Pre-test accommodation was a reliable covariate of accommodation [$F(1,11) = 12.98$, $p < 0.01$], while dark focus was not; this suggests that accommodation was not greatly affected by target position (i.e., 35 cm) relative to the various individual dark-focus positions. Perhaps these line targets provided good accommodation cues, even though they were presented against a dark background and viewed monocularly.

There was also no indication that tasks had any effect on dark focus. A repeated-measures t -test showed no reliable difference between the dark-focus estimates taken before (2.15D) and those taken after (2.42D) task performance.

To test the possibility of a relationship between accommodation and perceived size, we obtained a Pearson product-moment correlation from the 72 accommodation measures and size-judgment errors. The 72 values included 6 means (pre-test, post-test, and the 4 experimental conditions) for each of the 12 subjects. The two variables were not reliably correlated ($r = 0.04$, $p = 0.76$).

ACKNOWLEDGMENT

This research was performed under contract F46920-77-C-0117, Air Force Office of Scientific Research (AFOSR), with Professor Stanley Roscoe as Principal Investigator and Dr. Alfred Fregly as the Scientific Monitor for AFOSR. Louis Corl, Donald Fahrenkrog, and John Cone programmed the computer and prepared the equipment, while Dr. Robert Hennessy provided scientific guidance and consultation and B. Jay Martin aided in data collection.

REFERENCES

- Alpern, M. The eye. In H. Davson, The eye, Vol. 3, New York Academic Press, 1969.
- Cornsweet, T. N., & Crane, H. D. Servo-controlled infrared optometer. Journal of the Optical Society of America, 1970, 60, 548-554.
- Gawron, V. J. Eye accommodation, personality, and autonomic balance (TR BEL-79-2/AFOSR-79-6). Las Cruces, NM: New Mexico State University, Behavioral Engineering Laboratory, November 1979.
- Hull, J. C., Gill, R. T., & Roscoe, S. N. Locus of the stimulus to visual accommodation: Where in the world or where in the eye? (BEL-79-5/AFOSR-79-9). Las Cruces, NM: New Mexico State University, Behavioral Engineering Laboratory, November 1979.
- Iavecchia, J. H., Iavecchia, H. P., & Roscoe, S. N. The moon illusion: Apparent size and visual accommodation distance (Eng Psy-78-4/AFOSR-78-3). Champaign, IL: University of Illinois, Department of Psychology, April 1978.
- Malmstrom, F. V. Effects of concurrent mental activity on static and dynamic accommodation responses. Unpublished doctoral dissertation, Oklahoma State University, Stillwater, Oklahoma, December 1978.
- Malstrom, F. V., Randle, R. J., & Weber, R. J. Perceptual tunnel or perceptual sphere? Accommodation and stress. Paper presented at the Rocky Mountain Psychological Association Meeting, Salt Lake City, Utah, May 1975.
- Randle, R. J., Roscoe, S. N., & Pettitt, J. Effects of accommodation and magnification on aimpoint estimation in a simulated landing task (NASA-TP-1635). Washington, D. C.: National Aeronautics and Space Administration, 1980.
- Roscoe, S. N. Ground-referenced visual orientation with imaging displays: Final report (TR Eng-Psy-79-4/AFOSR-79-4). Las Cruces, NM: New Mexico State University, Behavioral Engineering Laboratory, November 1979.
- Roscoe, S. N., Olzak, L., & Randle, R. Ground-referenced visual orientation with imaging displays. Monocular versus binocular accommodation and judgments of relative size. AGARD, Athens, 1976.
- Skeffington, A. The totality of vision. American Journal of Optometry, 1957, 34, 241-255.
- Simonelli, N. M., & Roscoe, S. N. Apparent size and visual accommodation under day and night conditions (Eng-Psy-79-3/AFOSR-79-3). Champaign, IL: University of Illinois, Department of Psychology, April 1979.

Visual Cues in the Simulation of Low Level Flight

Edward J. Rinalducci
School of Psychology
Georgia Institute of Technology
Atlanta, GA 30332

Elizabeth C. Martin
Air Force Human Resources
Laboratory
Williams AFB, AZ 85224

and

Thomas Longridge
Air Force Human Resources
Laboratory
Williams AFB, AZ 85224

Abstract

Visual cues used by pilots to maintain altitude in low level flight simulation were examined. In particular, terrain texture in the form of black vs. white topped inverted cones, the presence or absence of vertical development, and the effects of rate of motion on terrain features were investigated using pilots who varied in flying experience. Less experienced pilots demonstrated increases in their mean altitude and RMS deviation with an increase in airspeed or with an increase in airspeed combined with a lack of vertical development in terrain features. Experienced pilots, on the other hand, only showed increases in mean altitude and RMS deviation with an increase in airspeed. No differences were found between the all black and the white topped cones.

Important visual factors are involved in flight simulation and the training of pilots. The various tasks especially required of U.S. Air Force pilots which involve these visual factors include formation flying and aerial refueling, low level flight weapons delivery, and approach and landing. The main objective of the present investigation was to examine the depth/distance cues employed in low altitude flight. Dr. E. L. Martin (AFHRL/OT, Williams AFB, AZ) in a preliminary study of terrain features in the computer image generated display (CIG) of the Advanced Simulator for Pilot Training (ASPT) found that those features consisting of inverted tetrahedrons or cones with white tops showed some advantage over all black cones in the maintenance of altitude in simulated low level flight. This finding seemed to suggest that a texture cue (in the form of contrast was being employed by pilots in the simulator when using the cone terrain features. In addition there has been some evidence that vertical development of terrain features should also influence the depth/distance judgments in flight (Harker and Jones, 1980; Stenger, Zimmerlin, Thomas, and Braunstein, 1981). Therefore,

the white triangular tops of the inverted cones were taken from the cones and placed for all practical purposes, on the ground (OMSL) in the CIG display of the ASPT. Also, as differences in texture or surface detail should vary with rate of motion, it was decided that airspeed should be manipulated as this should influence altitude maintenance.

Method

Subjects. Thirteen subjects participated in the first part of the study. The subjects in the first phase of the investigation were all B-course pilots. The B-course pilots were relatively inexperienced pilots (average flying time of 779 hours and SD = 684) who have recently completed undergraduate pilot training and were learning to fly the F-16. In the second part of the study, 6 T-course pilots were employed. The T-course pilots were transitioning from other aircraft (e.g., F-4, F-111, etc.) to the F-16, and therefore had more flying experience (average flying time of 1864 hours and SD = 767). All pilots had just completed at least 4 hours of training in the ASPT, F-16 cockpit before participating in the study. Thus, a total of 19 subjects were employed in the entire investigation.

Apparatus. All subjects were run in the ASPT which consists of two fully instrumented cockpits with wraparound visual systems. One cockpit is configured as the A-10 and the other as the F-16. The visual system which employs CIG consists of seven 36-inch CRTs which provide a visual field of + or - 150 degrees horizontally and +110 to -40 degrees vertically. This system has been described in greater detail by Buckland, Monroe, and Mehrer (1980).

Independent Variables. The major factors which were varied consisted of terrain texture and air speed. Terrain texture cues were varied by using inverted tetrahedrons or cones which had either white tops or were all black and by employing white triangles on the ground. The cones were 35 feet high and had about a 10 foot base. The triangles had no height and also had a 10 foot base. The triangles essentially consisted of the tops or bases of the inverted cones. The cones provided vertical development, whereas this cue was removed in the case of the triangles. Both types of cones and the triangles were distributed with an average density or separation of 1500 feet. Thus, there were three terrain features all with the same distribution: black cones with white tops, all black cones, and white triangles on the ground. In addition, there were two airspeeds: 300 KIAS and 540 KIAS. The difference in airspeed provided a rate of texture motion which could produce differences in a depth cue such as motion parallax.

Altitude was to be maintained under all conditions at 200 feet AGL. Ground elevation was OMSL. the 200 foot altitude was chosen so as the various terrain features would be readily visible but still represent low level flight.

Experimental Design and Procedures. The experimental design used in this study was a 3 X 2 factorial design (3 terrain features X 2 airspeeds) with repeated measures on each variable producing a 3 X 2 within-subjects design. Each student pilot was run for a total of 15 trials (about 1 hour and 10 minutes in duration). The first three trials were practice trials and the student pilots were informed of their altitude throughout the trial in order to provide knowledge-of-results or feedback. However, during the last 12 trials they were only told how close they came to the prescribed altitude at the end of the trial. Three of the six conditions were chosen at random for the 3 practice trials, and during the following 12 trials the 2 trials for each of the 6 conditions were distributed randomly.

Each student pilot received both a verbal and a written briefing prior to participation in the experiment. The student pilot was informed that he was to fly a prescribed course at an altitude of 200 feet and to maintain an airspeed of either 300 or 540 KIAS. The course consisted of three legs arranged at right angles. Each leg was seven miles long and two miles wide. An aiming tower measuring 450 feet was positioned at the end of each leg. A tone was presented through the student pilot's headset and served as a cue to initiate a left or right turn as required. The turn was a 90 degree heading change.

Each pilot was initialized to the desired altitude at the beginning of each trial. In addition, whenever a pilot fell below or exceeded his prescribed airspeed by more than 40 KIAS an automated voice said "Low, Low", etc. or "High, High", etc. until the appropriate increases or decreases in airspeed were made.

All pilots were questioned immediately after participating in the experiment to determine their preference for the different terrain features (in rank order), and the different cues they used to maintain altitude.

Dependent Measures and Data Analysis. The ASPT has the capacity to record and store a wide variety of variables associated with low level flight simulation. However, the ones that were analyzed in this study included the mean altitude and RMS deviation (from 200 feet) for the total course, as well as the mean altitude and RMS deviation for the level flight portions of the course and the turns. In addition, the number of terrain crashes for each of the six conditions were recorded.

Results

Due to space limitations only the results in terms of mean altitude for the entire course will be considered. The other response measures (i.e., RMS deviation for the entire course as well as level flight and turns and the mean altitude for level flight and turns) provided the same outcome. MANOVAs

were carried out on the data, and Tukey tests were employed where appropriate.

In general, for the B-course pilots, it was found that mean altitude was most affected by the variables of airspeed ($F = 157.6789$; $df = 1, 24$; p less than 0.00001) and texture of terrain features ($F = 15.072$; $df = 2, 24$; p less than 0.00003). Tukey tests (p less than 0.01) showed that mean altitude increased significantly with an increase in airspeed and when white triangles were used in place of either type of cone (all black or white topped). In addition, there was a significant subjects effect ($F = 3.3996$; $df = 12, 24$; p less than 0.0007).

The results for the T-course or more experienced pilots indicated that the only significant effects were for airspeed ($F = 15.8053$; $df = 1, 10$; p less than 0.00056) and subjects ($F = 3.4305$; $df = 5, 10$; p less than 0.012) with texture being non-significant.

The number of terrain crashes were determined for each terrain feature-airspeed combination for both B-course and T-course pilots. In general, the overwhelming number of terrain crashes for B-course pilots (16 out of 19) and T-course pilots (4 out of 5) occurred for the conditions which employed the white triangles. For both groups most crashes occurred in the right turn (20 out of 24).

Pilot subjects were questioned immediately after participating in the experiment with regard to their preferences of the three terrain features and the cues they used in maintaining their altitude. In general, B-course pilots preferred white topped cones to all black cones, whereas the T-course pilots preferred the black to the white topped cones. For all subjects, the white triangles were preferred the least.

Few differences were observed between the comments of B-course and T-course pilots with regard to the kinds of cues they used to maintain altitude. Typical comments are summarized as follows: (1) the HUD was used as a visual reference relative to the aiming towers or the horizon; (2) the use of peripheral vision; (3) the necessity of having vertically developed terrain features; (4) the relative height of the terrain features and the aiming towers; (5) the rate of movement of the terrain features; (6) the size of the white triangles; and (7) the use of the VVI on the HUD to provide an indication of altitude.

Discussion

In general, the results obtained with both groups of subjects (B-course and T-course pilots) in the present study showed that airspeed had the greatest effect on the maintenance of altitude (including mean altitude over the entire course, in level flight, and in the turns, as well as, mean RMS deviation over the entire course, in level flight, and in the turns).

In other words, an increase in airspeed from 300 to 540 KIAS produced a significant increase in mean altitude and RMS deviation. However, less experienced pilots (B-course) also showed significantly increased altitude and RMS deviation measures at the higher airspeed when there was a lack of vertical development in the terrain features (white triangles vs. cone). This effect was not found with the more experienced pilots (T-course), which suggests that they are better able to utilize other cues in their visual environment in the absence of vertically developed terrain features. No significant effects were found when texture in the form of contrast was manipulated (all black vs. white topped cones) for either group. Thus, vertical development appears to be the most important terrain feature. This finding is emphasized when the data for terrain crashes are examined. As previously noted most of the crashes occurred when using the white triangles for both groups. This indicates that vertical development is also important to the more experienced as well as the less experienced pilots.

A significant subject effect with both Bcourse and T-course pilots may well be due to the differences in flying experience. Pilots in both groups tended to vary both in the number of hours flown and in the type of aircraft flown.

It was also noted that most crashes occurred when pilots attempted to make a right turn. This finding may be due to the difficulty many pilots experience in making right vs. left turns.

Finally, in that no significant differences were obtained with two different types of cones, the one that is most edge efficient in the CIG display would be the most appropriate. As the all black cones are more edge efficient due to hidden contours than are the white topped cones, the black cones should be preferred as terrain features.

References

- Buckland, G. H., Monroe, E. G., and Mehrer, K. L. Flight simulator runway visual textural cues for landing. AFHRL-TR-79-81, August 1980.
- Harker, G. S., and Jones, P. D. Depth perception in visual simulation. AFHRL-TR-80-19, August 1980.
- Stenger, A. J., Zimmerlin, T. A., Thomas, J. P., and Braunstein, M. Advanced computer image generation techniques exploiting perceptual characteristics. AFHRL-TR-80-61, February 1981.

Selection for Command: What Really Counts?

Frederick V. Malmstrom, Edward Batchelor, Jr.,
Christopher J. Reamy, and Richard M. Coffman

Institute of Safety and Systems Management
University of Southern California
Los Angeles, California 90007

Abstract

Three grade levels of Air Force officers were surveyed for their ratings of which items from the Officer Career Brief listing were important for selection to a command position. Results indicated very strong agreement among all groups of officers, especially on the items of Duty and Performance, Professional Military Education, and Advanced Academic Degree. Results are discussed as to the generalizability of this study to the officer promotion process.

Introduction

Senior commanders faced with the requirement to select a junior commander from a large pool of seemingly qualified officers find they are faced with a tedious and difficult task. A common procedure in the selection process is for the senior commander to assemble the personnel files of, say, 20 or 30 officers on his desk and begin a laborious elimination process. The personnel file may contain scores of items on which the applicant could be selected or eliminated, from the applicant's academic level to his or her marital status. It is apparent in the selection process that the commander could be overwhelmed by an embarrassment of information from critical to trivial items. Hence, it would be most convenient in the selection process for a commander if he could narrow the list to relevant items.

The officer promotion selection board procedure is, likewise, such a personnel selection process, and it would be useful if the promotion selection item process could also be used by a division or wing commander. Unfortunately, as Hendrix (1971) and others have indicated, the criteria on which the promotion boards select officers are not open for examination. A side effect of this confidentiality is that both local personnel and unit commanders are in the uncomfortable position of explaining to an officer why he or she was not selected for promotion, even though the local personnel or unit commander was not privy to the workings of the selection board.

Therefore, the purpose of this study is threefold. First, it is desired to develop a tool useful to, say, a division or wing commander in the selection of a junior commander. It

would be most convenient to pare the personnel file down to a list of, say half a dozen items which are critical to selection. "Throw away" items could easily be eliminated by computer listings. Second, it would be desirable to see how junior and mid-level officers perceive the command selection process compared to those who actually do the selecting. Is there, for example, a certain naivete among junior officers in their perceptions of what is "necessary" for either promotion or command selection? Third, the question arises as to whether the results of this study could be extended to a set of guidelines which a junior or mid-level officer could use if he or she aspires to a command position or further promotion. It is to the aspiring officer that this final question is addressed.

Method

Subjects. The subjects were a total of 30 Air Force officers, 10 each drawn from three separate sources. The first group, Commanders, was taken from two Strategic Air Command air divisions (2 brigadier generals, 8 colonels) who were either commanders or deputy commanders. These officers had either previously served, or else were likely to serve, on Air Force-wide officer promotion selection boards.

The second group of officers, the Fast Burners, were local unit commanders and operations officers (6 lieutenant colonels, 4 majors) identified by the air division commander and executive officer as being possible excellent wing commanders sometime in the future. These officers had one or more of the following distinguishing characteristics: Previous below-the-zone selection, selection for senior service school, or across-the-board highest block officer effectiveness ratings.

The third group of 10 officers, Control, consisted of local unit majors, captains, and lieutenants selected randomly from the local wing personnel roster. Four of these officers had no aeronautical rating.

Procedure. Survey packets were sent to all officers, complete with cover letter and instructions. The subjects were requested to rate, on a scale of from 1 to 99 the 22 items from the Officer Personnel Brief as to their perceived importance in selecting an officer for a responsible command position. The subject was free to, for example, rate all items equally if he wished. Ties were acceptable. Response was either through distribution or return mail. Responses from Groups I and II were 100%. However, three of the officers from Group III failed to complete the survey, and another three officers were randomly selected.

Materials. The survey kit consisted of a letter of introduction and an example copy of an Officer Personnel Brief. Attached was the survey answer sheet with 22 items listed, in

alphabetical order, from Academic Level (Recent) to Total Flying Time. The alphabetical listing of each item is shown in the left-hand column of Table 1.

Results

Figure 1 shows, by Group, the mean score of each command selection item. The arrangement of items on the abscissa is arbitrarily set to follow in descending order of rated importance by the Commanders (Group I). There were striking differences between the relative importance of the 22 items and indications that all officers tended to have general agreement as to which items were important for selection for command.

An analysis of variance of the data revealed a main effect only for the rating of command selection items, $F(21,567) = 17.12, p < .001$. The apparent interaction of effect of groups x items only approached an acceptable level of significance ($p = .076$); therefore, this effect is questionable. However, as Figure 1 indicates, there is exceptional agreement between all officers in items 1 through 7, and for this reason the items are shown by groups.

A post-hoc Newman-Keuls test was performed to show the significance of each item relative to all other items. The results indicate that items 1 through 7 are more likely to differ from all other items. In particular, the most recent duty and performance section (which includes an abbreviated officer effectiveness rating index) carries the preponderance of weight in the selection process.

Discussion

The consistency with which the rating of the first seven items are perceived as a necessity for selection for command is most striking. The only point which seems to run contrary to the trend is the relatively low scoring of Flying Data (Recent) by the Control group. Perhaps this is because nearly half of the Control group had no aeronautical rating.

The high rating of Professional Military Education (Squadron Officer School, Air War College, etc.) and Method Obtained is worthy of comment. The completion of these programs takes considerably less time, expense, and effort than, say, a graduate degree, regardless of the method by which one completes the program. For example, for an officer to complete the Air Command and Staff Course by correspondence may, optimistically, require three months' worth of night and weekend study. By comparison, an external master's degree usually requires between one to two years to complete, not to mention books and tuition expenses. In other words, an advanced academic degree does seem to be important in the selection for command but does not seem to be as important as the professional military education programs.

Next, an analysis of the items which did not seem to have

Table 1. Newman-Keuls item analysis. Ratings of items desired for selection for command.

	Duty & Perf (Recent)	Duty & Perf (2d Recent)	Duty & Perf (3d Recent)	PME & Meth (Recent)	PME & Meth (2d Recent)	Fly Data (Recent)	Academic (Recent)	Perm Grade /DOR	Temporary Grade	Fly Data (2d Recent)	TAFSCOM Date	Previous Grade B2	Marital Status	Current Aero Rate	Date Arr Station	Academic (2d Recent)	Duty AFSC	Primary AFSC	Total Flying Time	Last Overseas	Source of Commission	Combat Miss (RVN)
1. Duty & Performance (Recent) (g)	11.8	15.2	15.9	24.1	24.2	27.2	27.9	29.3	30.0	34.4*	34.7*	39.1*	39.1**	39.2**	39.3**	41.0**	45.8**	45.8**	46.2**	59.4**	60.6**	65.3**
2. Duty & Performance (2nd Recent) (h)	-	3.4	4.1	12.3	12.4	15.4	16.1	17.5	18.2	22.6	22.9	27.3	27.3	27.4	27.5	29.2	34.0	37.4*	47.6**	48.8**	53.3**	
3. Duty & Performance (3rd Recent) (i)	-	-	.7	8.9	9.0	12.0	12.7	14.1	14.8	19.2	19.5	23.9	23.9	24.0	24.1	25.8	30.6	31.0	44.1**	45.4**	50.1**	
4. PME & Method Obtained (Recent) (p)	-	-	-	8.2	8.3	11.3	12.0	13.4	14.1	18.5	18.8	23.2	23.2	23.3	23.4	25.1	29.9	30.3	43.5*	44.7**	49.4**	
5. PME & Method Obtained (2nd Recent) (q)	-	-	-	-	.1	3.1	3.8	5.2	5.9	10.3	10.6	15.0	15.0	15.1	15.2	16.9	21.7	22.1	35.3*	36.5*	41.2**	
6. Flying Data (Recent) (j)	-	-	-	-	3.0	3.7	5.1	5.8	10.2	10.5	14.9	14.9	15.0	15.1	16.8	21.6	22.0	35.2	36.4*	41.1**		
7. Academic Degree (Recent) (a)	-	-	-	-	.7	2.1	2.8	7.2	7.5	11.9	11.9	12.0	12.1	13.8	18.6	19.0	32.2	33.4	38.1*			
8. Permanent Grade/Date of Rank (o)	-	-	-	-	1.4	3.1	6.5	6.8	11.2	11.2	11.3	11.4	13.1	20.9	18.3	31.5	32.7	37.4				
9. Temporary Grade (u)	-	-	-	-	-	.7	5.1	5.4	9.8	9.6	9.9	10.0	11.7	16.5	16.9	30.1	31.3	36.0				
10. Flying Data (2nd Recent) (k)	-	-	-	-	4.4	4.7	9.1	9.1	9.2	9.3	11.0	15.8	16.2	29.4	30.6	35.3						
11. Total AF Service Commission Date (t)	-	-	-	-	.3	4.7	4.7	4.8	4.9	6.6	11.4	11.8	25.0	26.2	30.9							
12. Previous Grade Below the Zone (r)	-	-	-	-	4.4	4.4	4.5	4.6	6.3	11.1	11.5	24.7	25.9	30.6								
13. Marital Status (m)	-	-	-	-	.0	.1	.2	1.9	6.7	7.1	20.3	21.5	26.2									
14. Current Aeronautical Rating (d)	-	-	-	-	.1	.2	1.9	6.7	7.1	20.3	21.5	26.2										
15. Date Arrived Station (e)	-	-	-	-	.1	1.8	6.6	7.0	20.2	21.4	26.1											
16. Academic Degree (2nd Recent) (b)	-	-	-	-	1.7	6.5	6.9	20.1	21.3	26.0												
17. Duty AFSC (f)	-	-	-	-	4.0	5.2	18.3	19.6	24.3													
18. Primary AFSC (n)	-	-	-	-	.4	13.6	14.8	19.5														
19. Total Flying Time (v)	-	-	-	-	13.2	14.4	19.1															
20. Last Overseas Date (l)	-	-	-	-	1.2	5.9																
21. Source of Commission (s)	-	-	-	-	4.7																	
22. Combat Missions (RVN only) (c)	-	-	-	-	-																	

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

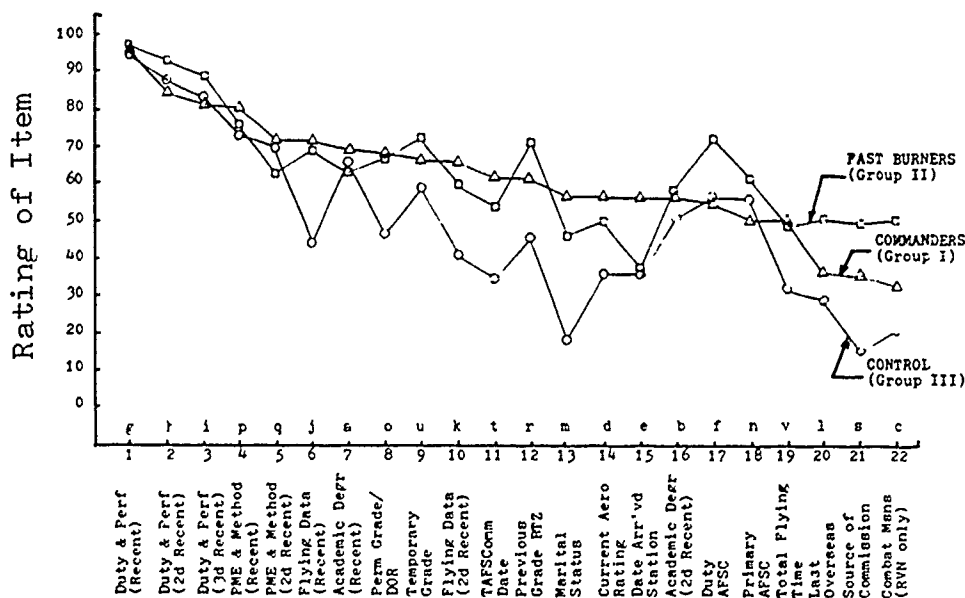


Figure 1. By group ratings of items desired for selection for command.

relative importance is in order. The Newman-Keuls test indicates that items 8 through 22 did not differ from each other statistically, and might possibly, be regarded as unimportant or "throw-away" items in the command selection process. Even within the list of low-rated items, there is a hierarchy which warrants comment. Marital Status is rated rather high (13th), but largely because it is held in high regard by the older, higher ranking Commanders. Presumably, therefore, most of Group I have well-established marriages. Moreover, Marital Status rates considerably higher than Source of Commission. Thus, an officer would do his or her career better by having a marriage certificate than a diploma from a service academy! Perhaps most curious is the observation that Combat Missions ranks dead last as a criterion for selection for command. It would seem, therefore, that combat experience counts for very little -- even when one is being considered to command a combat unit.

The extent to which the results of this survey can be generalized to apply to other selection techniques, particularly the officer promotion selection boards is, of course, both unknown and tantalizing. Because selection boards do not publish their criteria for promotion selection, we can only speculate. We feel, however, that the correlations between this survey and the boards are bound to be fairly strong for the following reasons: First, Group I was composed of officers who were likely to be on such a board. Second, it is generally the overall quality of fitness to command that officers are promoted; technical expertise and job knowledge are not sufficient promotion criteria. Third, the extreme high rating of a recent officer effectiveness rating seems to support the common agreement among officers that a recent unfavorable OER is sufficient to override all other considerations for promotion. Conversely, it would seem that possession of high ratings on the first seven items would be sufficient to overcome weaknesses in all other areas.

The overall question of whether this selection technique (or even a promotion board) method is valid is, of course, not answered. [For a comprehensive treatment of the validity of selection systems, see Dunnette & Borman (1979).] The only true validation of a selection technique would be observation of these officers' performance in actual combat conditions, a situation which will, hopefully, never occur.

Bibliography

Dunnette, M. & Borman, W. Personnel selection and classification systems. In Annual Review of Psychology, 1979, 30, 477-525.

Hendrix, W. Strategic Air Command Minuteman personnel selection study. Published in Proceedings of the 1971 Psychology in the Air Force Symposium, 20-22 April 1971. Department of Life and Behavioral Sciences, USAF Academy, Colorado 80840.

Thanks to Donna L. Miller, Informatics, Inc. for assistance in data analyses.

Development and Implementation of the Officer Training School Selection Equation

Lynn M. Scott
Air Force Human Resources Laboratory, Brooks AFB TX

Abstract

The application of the Policy Specifying technique was used to develop a selection equation for Officer Training School. The composite policy of 7 Air Force Agency Representatives was modeled mathematically and produces a zero to 100 point rating that is the synthesis of values from 14 variables. Internal validity tests using both simulated and actual case data revealed that the equation is an accurate model of Air Force policy, and that the equation can successfully identify high risk applicants.

The Air Force Officer Training School (OTS) is the component of the precommissioning program triad tasked with short lead time production of junior officers. As with the Air Force Academy and the Air Force Reserve Officer Training Corps (ROTC), selection of the most qualified applicants is paramount. The Air Force Academy uses a regression model score in conjunction with an officer panel rating to select qualified candidates. Air Force ROTC also uses a regression policy model in its Weighted Professional Officer Course Selection System (WPSS). Unlike these two programs, however, OTS places a large dependence on a selection panel composed of volunteer officers to evaluate most applicants.

The number of applications to the OTS has increased considerably over the past three years. Accompanying this growth has been Air Force Recruiting Service's concern about its selection accuracy. A recruiting service study of selection panel scoring revealed inconsistencies between panel members on the same panel and between selection panels conducted at different times. The cause of the inconsistencies was speculated to be an absence of enforceable scoring standards and stimulus overload. Despite instructions in the scoring procedure and scoring rehearsals, panel scoring is still plagued with leniency and/or central tendency effects. Moreover, the conclusions of Cornelius and Lyness (1980) concerning stimulus overload suggests that panel members may be overloaded when they examine and weigh a lengthy list of applicant data (i.e., aptitude test scores; grade point average; academic major; demonstration of communications skills; civilian work record; military performance; reported skills, hobbies, and desire to travel; and subjective assessments of character, motivation and leadership from third parties). Considering all this information, Air Training Command and the Manpower and Personnel Center jointly requested the research and development of an automated selection process/selection advisory for Officer Training School; which is described in this paper.

Method

The policy specifying technique (Ward, 1977; Ward, Pina, Fast & Roberts, 1979) was used to create the selection algorithm. Policy specifying is a method of combining many different types of information into a single indicator of utility or payoff. The procedure involves identification of the policy expert (or group of policy experts) who will interact with the policy modeling specialist; selection of the basic variables which are combined to produce a single indicator of utility; specification of a pairwise hierarchical relationship among all of the variables; and, finally specification of each pairwise function in the hierarchy. This process allows the policy experts to combine many variables in complex non-linear and interactive ways to produce a single indicator of utility.

Participants

Representatives at the Air Force policy interpretation and implementation level were selected from the following organizations: USAF Recruiting Service; Air Training Command (ATC) Officer Commissioning and Management; ATC Training Technology Division, ATC Technical Training; ATC Education Plans and Programs Directorate; the Officer Procurement Branch at the Manpower and Personnel Center and the Manpower and Personnel Division at the Air Force Human Resources Laboratory. The total specifying working group membership was 14 officers.

Variable Selection and Hierarchy Development

A "degree of importance" questionnaire containing 22 existing and potential selection variables was constructed to evaluate the variables. The questionnaire was completed by 47 officers representing Air Training Command Headquarters, Headquarters Recruiting Service, the Manpower and Personnel Center and members of an Airman Education and Commissioning Program (AECF) selection board. The rating reliability for a single rater, r_{11} , was .20 and the reliability for this group of raters, r_{kk} , was .92. The variables with the 10 highest mean ratings were: active duty experience, indication of reading ability; indication of writing ability; education background (tie); participation in activities requiring leadership (tie); college GPA; Air Force Officer Qualifying Test (AFOQT) qualitative subtest score; AFOQT verbal subtest score; participation in activities requiring management of resources and; indication of aptitude for military life. Apart from the questionnaire results, each variable's practicality was considered. Consequently, candidate variables such as indicators of writing and reading ability were not selected and variables measuring college quality (the Patton Rating¹), recruiter evaluation, extracurricular activities, and non-military awards were included in the final selection of 14 variables.

¹The Patton Rating was also developed by policy specifying. The policy incorporates four variables; freshman Scholastic Aptitude Test (SAT) average, freshman class selection ratio, the percentage of faculty members with Doctorate or professional degrees, and the percentage of graduate students at the institution. Description of the research and development supporting the Patton Rating is beyond the scope of this paper.

The variable hierarchy that was developed by the policy specifying working group may be interpreted as a framework for the pairwise organization of the 14 variables into three domains; an objective cognitive measures profile, a military performance profile, and a civilian performance profile. The selection hierarchy is shown in Figure 1.

Payoff Metric

The payoff metric was a 100 point behavioral expectation scale (Smith & Kendall, 1963). Every five point increment was anchored with an expected behavior in either OTS or an active duty officer. Anchors were developed using the procedure detailed in Campbell, Dunnette, Arvey and Hellervik (1973).

Procedure

The working group articulated and discussed selection policy in weekly three hour sessions. A consensus of "payoff" to the Air Force of value combinations for the 14 paired variables was the goal of each meeting. Throughout the modeling process, the group received feedback on their policy's effects. Feedback included distributions and equivalency tables of payoff award; examining the results of the policy algorithm's processing of simulated records and; examining the independent effects of each selection variable on the final selection score.

When the complete policy had been modeled it was internally validated through two mock selection board proceedings. In the first mock board (hereafter mock board I), 50 applicants' records were randomly selected from a pool of records scored earlier in the week by an actual OTS selection panel. The second mock board (hereafter mock board II) validation used 50 randomly selected applicant folders from the Recruiting Service declination and rejection files. This board was convened to establish an alternate form reliability for the policy and the judges response patterns. Although the records were selected from declination and rejection files, they still represented potentially successful and unsuccessful applicant records. The working group members served as judges in both mock boards. Evaluation of the algorithm's efficacy consisted of examining the convergent validity of (1) correlations between the algorithm's applicant ranking and the mock board's applicant ranking based upon an average rating; (2) the correlations between the algorithm's applicant ranking and the actual OTS panel's ranking; (3) the correlation of the mock board's applicant rankings with the OTS panel's rankings and; (4) calculation of the mock board's interrater reliability. The effectiveness of the behavioral expectation scale was assessed by comparing mock board means with algorithm means. This kind of comparison was necessary because the entire selection policy was specified through the same behavioral expectation metric used by the mock board members. Additional comparisons were made between military and civilian groups in each mock board. These included evaluation mean board score differences, mean algorithm score differences, and comparing subgroups' Pearson correlations between mock board and algorithm scores.

Results

The board members employed the behavioral expectation scale effectively. The mean rating for mock board I was 59.04 and the standard deviation was 14.03. Comparing this mean with the algorithm mean ($M = 61.898$, $\sigma = 21.10$)

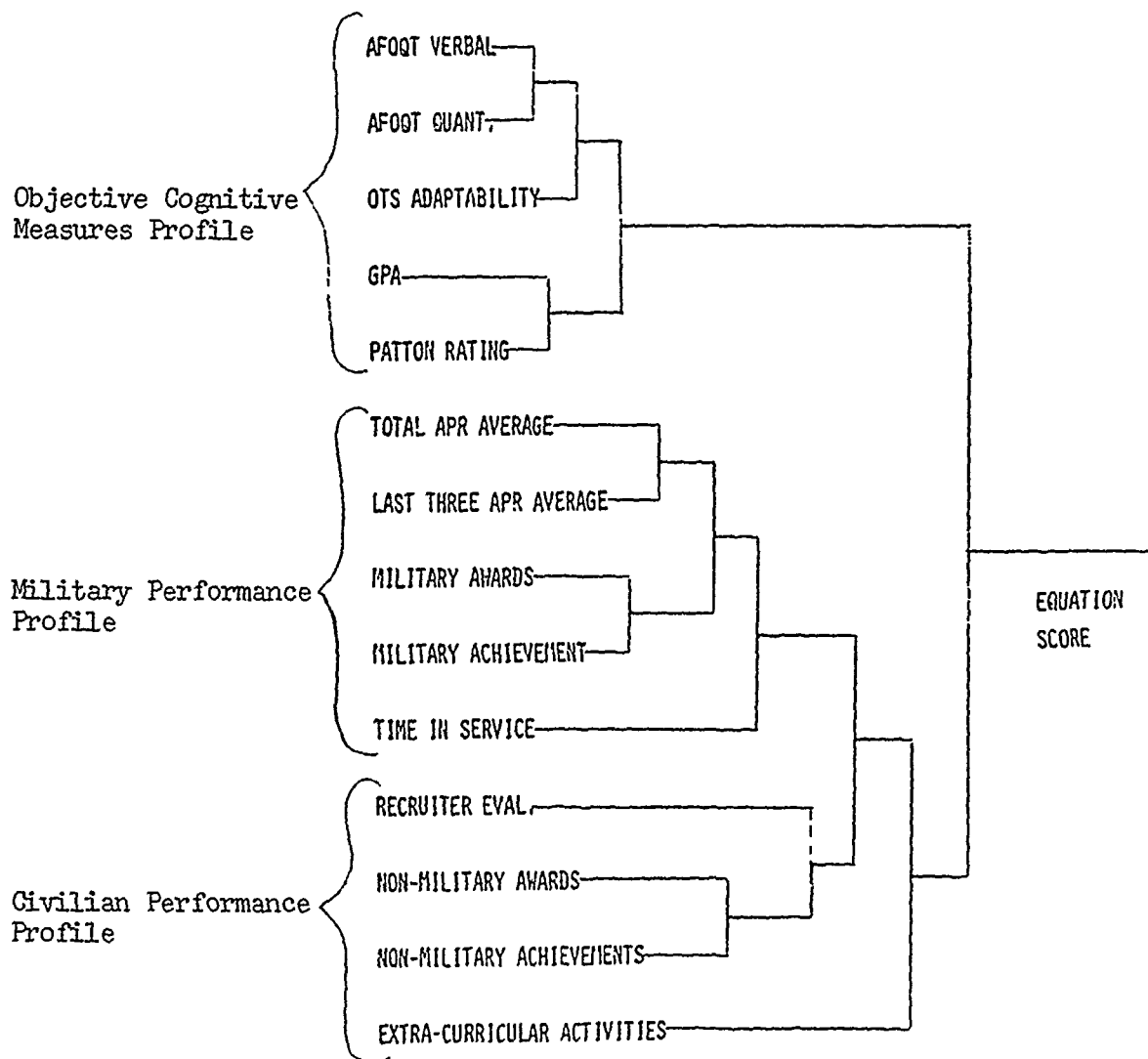


Figure 1. The OTS Selection Hierarchy

for the same records did not yield a significant difference ($t(49) = .22$, n.s.). The same check was made for mock board II data. This board's mean rating was 55.43 and its standard deviation was 13.01. The t test comparing the algorithm's statistics ($M = 61.866$, $\sigma = 18.04$) was also not significant ($t(49) = .57$, n.s.). A Pearson correlation was computed between the mock board mean scores and the algorithm scores for both boards. The mock board I correlation was .46 ($p \leq .01$) and for mock board II, .37 ($p \leq .01$). The intraclass correlation r_{11} (Winer, 1971) for mock boards I and II was .68 and .65 respectively.

Rank order correlations were computed between the three scoring systems: the actual OTS panel rankings; the average mock board rankings; and the algorithm rankings. The correlation coefficient for mock board rankings and OTS panel rankings was .55 and .52 for the mock board rankings and algorithm ranking comparisons. The rank order correlation between the OTS panel and the algorithm was .37. Mock board and algorithm scores were found to produce fewer rank ties than the OTS panel scores.

An inspection of scatter pilots for military and civilian subgroups revealed a tighter pattern of dispersion for civilian subgroups in both mock boards. Subgroup mean board score differences within each mock board were not significant, ($t_{MBI}(48) = .05$, n.s.; $t_{MBII}(47) = .35$, n.s.); nor were the differences between mean algorithm subgroup scores ($t_{MBI}(48) = .5$, n.s.; $t_{MBII}(47) = .32$, n.s.). For each mock board, a Pearson r was also computed between board scores and algorithm scores within each subgroup. Mock board I coefficients were .54 and .57 for the military and civilian subgroups respectively. Correlations in mock board II were .38 for the military subgroup and .37 for the civilian subgroup. All coefficients were significant at the $p \leq .01$ significance level.

Discussion

An important concern throughout this project was whether the mathematical model being created really represented the working group's policy. Furthermore, there existed a possibility that the policy expressed at the development table was totally different from the policy employed at the mock board table. The findings of this study should alleviate those concerns. The algorithm accurately models Air Force selection policy, identifies low potential applicants and produces a superior rank ordering of applicants. The behavioral expectation scale, the translator of policy, produced the kinds of reliabilities reported in much of the current literature (Jacobs, Kafry & Zendick, 1980).

The differences between characteristics of military and civilian scores are probably a result of fewer selection variable measures for the civilians than for the military. It would probably be difficult to correct this discrepancy since gaining access to civilian work records is not feasible. Therefore, greater scrutiny has to be placed in the recruiting officer's evaluation of the applicant and a listing of awards, achievements and extracurricular activities. The planned implementation of the behavioral expectation scale as a recruiting officer's rating tool and using the restricted awards, achievements and extracurricular activities lists should increase the civilian score dispersions.

In all, the project was judged a success. The selection system's implementation plan has begun at Air Force Recruiting Service Headquarters and AFHRL has begun a study to assess its external validity. The system appears to offer great promise in selecting future Air Force officers.

References

- Campbell, J.P., Dunnette, M.D., Arvey, R.D., & Hellervik, L.V. The development and evaluation of behaviorally based rating scales. Journal of Applied Psychology, 1973, 57, 15-22.
- Cornelius, E.T., & Lyness, K.S. A comparison of holistic and decomposed judgement strategies in job analysis by job incumbents. Journal of Applied Psychology, 1980, 65, 155-163.
- Jacobs, R., Kafry, D., & Zedeck, S. Expectations of behaviorally anchored rating scales. Personnel Psychology, 1980, 33, 595-640.
- Smith, P.C., & Kendall, L.M. Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 1963, 47, 149-155.
- Ward, J.H., Jr. Creating mathematical models of judgement processes: From policy capturing to policy specifying. AFHRL-TR-77-47, AD-A048 983. Brooks AFB, TX: Occupational and Manpower Research Division, Air Force Human Resources Laboratory, August 1977.
- Ward, J.H., Jr., Pina, M., Jr., Fast, J.C., & Roberts, D.K. Policy specifying with application to personnel classification and assignment. Paper presented at the meeting of the Military Testing Association, San Diego, October 1979.
- Winer, B.J. Statistical principles in experimental design. New York: McGraw-Hill, 1971.

NONTRADITIONAL ADMISSIONS FACTORS:
THE SPECIAL 100 PROJECT

Harald E. Jensen and William J. Strickland
Directorate of Institutional Research
USAF Academy, Colorado 80840

Abstract

Approximately 2,000 USAF Academy candidates were rated using the Candidate Evaluation Form. The items on the form were hypothesized to be associated with retention at the Academy and dealt with a variety of characteristics, such as indication of early/long-term interest in attending USAFA, military family background, college activities, etc. After the Class of 1984 had been filled through the normal admissions process, an additional 100 offers of appointment were made to the top scorers on the Candidate Evaluation Form. For the entire sample, regression analyses showed a fairly low correlation between the items on the form and retention, but the correlations rose considerably when the sample was segmented into ranges of academic ability. Other analyses are discussed and future research is addressed.

NONTRADITIONAL ADMISSIONS FACTORS:

THE SPECIAL 100 PROJECT

BACKGROUND

Historically, cadet attrition at the Air Force Academy has been similar to that of Annapolis and West Point. Typically, attrition at Annapolis was a little lower and that of West Point a little higher than that of the Air Force Academy. Then, about 1979, retention at Annapolis improved. From 35.4% for the Class of 1978, the attrition rate fell to 29.8% for the Class of 1979, then to 26.4% for the Class of 1980. Attention had always been given to the Air Force Academy's attrition rate, but this comparison increased the focus, particularly in light of attrition for the current classes at each academy:

	<u>1982</u>	<u>1983</u>	<u>1984</u>	<u>1985</u>
USAFA	41.3%	33%	22.1%	11.3%
USMA	35.4%	33.6%	22.8%	13.7%
USNA	22.2%	20.3%	10.5%	6.7%

In November 1979, Lt General K. L. Tallman, USAFA Superintendent, initiated a project with the objective of selecting a group of candidates for the Class of 1984 who would be particularly motivated to attend and graduate from the Air Force Academy. To identify this group of candidates, an experimental Candidate Evaluation Form was developed which tapped information not explicitly used in the admissions process, but hypothesized to be associated with retention at the Air Force Academy.

OBJECTIVES

The primary objective of the research was to examine candidate information which had not been explicitly used in the admissions process to see if we could increase our ability to predict retention at the Air Force Academy. If some of this information proved useful, the next step would be to incorporate it into the normal admissions process.

OVERVIEW

Approximately 2,000 candidate folders were rated using the Candidate Evaluation Form.* The items on the form dealt with a variety of characteristics, such as whether the student attended more than one high school, if the student was a "late bloomer," indication of early/long-term interest in attending USAFA, military family background, etc. Points were assigned to each item on the form and the candidate was given those points if he or she possessed that particular characteristic. The points for each candidate were then totaled and the candidates were rank ordered. Once the Class of 1984 had been filled, the top 100 scorers on the Candidate Evaluation Form were offered appointments. Of these 100, 87 accepted

* A copy of the Candidate Evaluation Form and a description of the items can be obtained by contacting: Harald E. Jensen, USAFA/RRE, USAF Academy CO 80840. Autovon 259-3225.

the appointment and reported for Basic Cadet Training. This group then became known as the "Special 100." Regression analyses were then performed to determine how well the Candidate Evaluation Form can predict retention, cumulative grade point average (GPA), and Military Performance Average (MPA).

ANALYSIS

As of 1 January 1982, the attrition rate for the Special 100 was 21.8%, slightly lower than the 22.7% rate for the rest of the Class of 1984. Table 1 shows the regression analysis which was performed between the items on the Candidate Evaluation Form, the Academic Composite, the Liaison Officer's Recommendation, and cumulative Fourth Class GPA. The multiple correlation of .601 is quite high, but the Academic Composite is responsible for most of that. Very little predictive validity was added when the items on the Candidate Evaluation Form were introduced.

Looking at the prediction of retention (.188) in Table 2, we get just about the same general picture. Earlier studies have shown that the Academic Composite and the Liaison Officer's overall recommendation are the two best predictors of retention. With this sample, this holds true, with the addition of the evaluation form's items not increasing our predictive validity by very much. Regressing the Candidate Evaluation Form against retention and excluding Academic Composite and the Liaison Officer's Recommendation from the equation, the multiple correlation was .15.

In Table 3, prediction of Fourth Class cumulative MPA using the Candidate Evaluation Form, Academic Composite, and Liaison Officer's Recommendation, reveals results similar to predicting GPA. Inclusion of all predictors yielded a multiple correlation of .304, and, once again, the greatest proportion of predictive ability comes from the Academic Composite. Excluding both Academic Composite and Liaison Officer's Recommendation showed the multiple correlation dropping to .190.

In Table 4, the total sample was broken down into approximate thirds based upon Academic Composite score: (1) 2960 and below, (2) Greater than 2960 and below 3252, and (3) Greater than 3252. This was done to determine whether predictive validity was approximately the same for several levels of academic ability. Rather interestingly, the multiple correlation for the sample with 2960 and below was .284, and then took a substantial jump to .386 and .382 for the next two-thirds of the distribution of Academic Composite scores. There was very little commonality of items appearing in these three regression equations and, surprisingly, all but one of the first six variables entering the equation for the sample with an Academic Composite greater than 3252 were negatively correlated with retention. This presents problems in a selection sense in that, contrary to the initial hypothesis that possession of the characteristics on the form would lead to greater retention, with this group, possession of the characteristics indicated greater attrition. The same jump in the magnitude of the correlation was evident when the Candidate Evaluation Form was regressed against Cumulative Fourth Class MPA. The multiple correlations for the lower, middle, and upper thirds of the distribution of Academic Composite scores were .287, .386, and .395.

When we cut the distribution of Academic Composite scores at 3252 and looked at those below this level, the multiple correlation was .237 for predicting retention.

We evaluated the *a priori* scoring system which placed a specific value for each item, which, in turn, was used to compute five separate scores: Academic Adjustment, Motivation Adjustment, Military Family, Unusual Leadership, and Total Score. Table 5 presents the results obtained when we regressed these scores against retention. The multiple correlation was a very low .046 and when Academic Composite score and the Liaison Officer Recommendation were added, only .116 was achieved. From this, it can be seen that the original scoring scheme somewhat lacks the precision of optimal weighting.

SUMMARY AND CONCLUSIONS

Preliminary results from the analyses performed using the Candidate Evaluation Form indicate that with one year of attrition data, the information on the form does not increase our ability to predict retention very much beyond that achieved by the Academic Composite score and the Liaison Officer's Recommendation. When we look at cadets falling within separate segments of the distribution of Academic Composite scores (i.e. below 2960, between 2960 and 3252, and above 3252), we get a different picture. The ability to predict retention rises considerably within each segment, increasing from .188 for the entire sample, .284 for those equal to or below 2960, .386 for those greater than 2960 and below 3252, and .382 for those above 3252.

Looking at other criteria, such as Cumulative Fourth Class MPA and GPA, the Candidate Evaluation Form again does not add much predictive validity beyond the current use of the Academic Composite and the Liaison Officer's Recommendation.

The current method of assigning points to each of the items on the Candidate Evaluation Form was not supported by the data. It appears that the weights should be changed and perhaps simple unit weighting throughout would be more appropriate.

In view of the initial results, no recommendations can be made to change the information currently used in the admissions process. However, a final conclusion at this time would be premature; a more definitive picture of the Candidate Evaluation Form's utility will come in August 1982 when the Class of 1984 reaches its active service commitment point and attrition begins to stabilize.

FUTURE RESEARCH

We will continue to track the attrition and performance of the Class of 1984. Additionally, the Candidate Evaluation Form was revised for the Class of 1985 and a clearer set of guidelines was formed for defining each of the items on the form. Lack of standardization in completing the Candidate Evaluation Form was a major problem with the Class of 1984 and these guidelines should help considerably. Nearly 3,000 candidate folders for the Class of 1985 were evaluated and this data will be statistically analyzed at the end of the Fourth Class year.

Another area in which the Candidate Evaluation Form will be examined is the identification of those candidates who are likely to decline an Air Force Academy appointment. It may be that some of the information on the Candidate Evaluation Form, together with academic information and the Liaison Officer's Recommendation, may help admissions personnel identify those most likely to accept or decline an appointment and, thus, more precisely estimate the number of offers needed to fill a particular class.

TABLE 1

PREDICTION OF CUMULATIVE
4th CLASS GRADE POINT AVERAGE ^{1/}
(N = 688)

<u>VARIABLE</u>	<u>ZERO-ORDER CORRELATION</u>	<u>MULTIPLE CORRELATION</u>
ACADEMIC COMPOSITE	.571	.571
UNUSUAL NONATHLETIC LEADERSHIP	-.090	.578
ACTIVE DUTY OFFICER PARENT	.101	.582
RETIRED OFFICER PARENT	.058	.585
ACTIVE DUTY ENLISTED PARENT	-.036	.588
STRONG INTEREST AND PERFORMANCE IN MATH AND SCIENCE	.166	.590
FINAL MULTIPLE CORRELATION (27 VARIABLES)		<u>.601</u>

TABLE 2

PREDICTION OF RETENTION ^{1/}
(N = 922)

<u>VARIABLE</u>	<u>ZERO-ORDER CORRELATION</u>	<u>MULTIPLE CORRELATION</u>
ACADEMIC COMPOSITE	.07	.073
LO'S OVERALL RECOMMENDATION**	.07	.096
RETIRED OFFICER PARENT*	.05	.109
RENOMINEE*	-.06	.119
STRONG COLLEGE RECORD WITH LOW PAR*	.04	.134
ACTIVE DUTY OFFICER PARENT*	.04	.142
FINAL MULTIPLE CORRELATION (27 VARIABLES)		<u>.188</u>

^{1/}Tables show the first 5 or 6 variables which entered the stepwise regression equation. Beyond these variables, additional entries accounted for very little variance and are not shown.

* Coding for these variables was "1" if the characteristic was present; "0" otherwise.

** LO's recommendation ranges from "1" (Not Recommended) to "5" (Highly Recommended).

TABLE 3

PREDICTION OF CUMULATIVE
4th CLASS MILITARY PERFORMANCE AVERAGE
(N = 763)

<u>VARIABLE</u>	<u>ZERO-ORDER CORRELATION</u>	<u>MULTIPLE CORRELATION</u>
ACADEMIC COMPCSITE	.205	.205
LO RECOMMENDATION	.157	.246
EARLY/LONG-TERM INTEREST	.115	.265
UNUSUALLY STRONG ATHLETIC LEADERSHIP	.025	.272
UNUSUAL NONATHLETIC LEADERSHIP	.072	.277
ACTIVE DUTY OFFICER PARENT	.044	.280
FINAL MULTIPLE CORRELATION (27 VARIABLES)		<u>.304</u>

TABLE 4

PREDICTION OF RETENTION BY ACADEMIC COMPOSITE GROUPING

ACADEMIC COMPOSITE EQUAL TO OR
LESS THAN 2960 (N = 270)

<u>VARIABLE</u>	<u>ZERO-ORDER CORRELATION</u>	<u>MULTIPLE CORRELATION</u>
STRONG COLLEGE RECORD	.11	.110
UNUSUAL INPUT FROM CANDIDATE	-.05	.149
RETIRED OFFICER PARENT	.11	.190
LO EVALUATION	.08	.206
PREP SCHOOL ATTENDANCE	.05	.218
ACTIVE DUTY OFFICER PARENT	.07	.229
FINAL MULTIPLE CORRELATION (27 VARIABLES)		<u>.284</u>

ACADEMIC COMPOSITE GREATER
THAN 2960 AND LESS THAN 3252 (N = 269)

<u>VARIABLE</u>	<u>ZERO-ORDER CORRELATION</u>	<u>MULTIPLE CORRELATION</u>
LO EVALUATION	.13	.130
STRONG INTEREST AND PERFORMANCE IN MATH AND SCIENCE	-.13	.184
EARLY/LONG-TERM INTEREST	.12	.222
CIVIL AIR PATROL	.12	.258
PREP SCHOOL ATTENDANCE	.07	.284
UNUSUAL ACHIEVEMENT IN NON-SCHOOL ACTIVITIES SUCH AS WORK	.08	.304
RENOMINEE	-.09	.318
FINAL MULTIPLE CORRELATION (27 VARIABLES)		<u>.386</u>

TABLE 4 (Continued)

ACADEMIC COMPOSITE GREATER
THAN 3252 (N = 285)

<u>VARIABLE</u>	<u>ZERO-ORDER CORRELATION</u>	<u>MULTIPLE CORRELATION</u>
UNUSUAL ACHIEVEMENT IN NONSCHOOL ACTIVITIES SUCH AS WORK	-.18	.180
UNUSUAL NONATHLETIC LEADERSHIP	-.15	.227
AFJROTC/AFROTC	-.13	.265
COLLEGE ACTIVITIES	-.16	.291
UNUSUAL LETTERS OF RECOMMENDATION	-.12	.313
VISITS TO AFA	.07	.327
FINAL MULTIPLE CORRELATION (27 VARIABLES)		<u>.382</u>

ACADEMIC COMPOSITE LESS
THAN 3252 (N = 538)

<u>VARIABLE</u>	<u>ZERO-ORDER CORRELATION</u>	<u>MULTIPLE CORRELATION</u>
LO EVALUATION	.10	.104
RETIRED OFFICER PARENT	.08	.127
UNUSUAL NONATHLETIC LEADERSHIP	.07	.144
STRONG INTEREST AND PERFORMANCE IN MATH AND SCIENCE	-.06	.160
STRONG COLLEGE RECORD	.08	.174
FINAL MULTIPLE CORRELATION (26 VARIABLES)		<u>.237</u>

TABLE 5

PREDICTION OF RETENTION USING PART AND TOTAL SCORES
FROM THE CANDIDATE EVALUATION FORM (N = 922)

	<u>ZERO-ORDER CORRELATION</u>
ACADEMIC ADJUSTMENT	-.006
MOTIVATION ADJUSTMENT	-.014
MILITARY FAMILY	.041
UNUSUAL LEADERSHIP	-.009
TOTAL SCORE	.008
FINAL MULTIPLE CORRELATION	<u>.046</u>

Norman J. Kerr, Harry L. Bowman

Characteristics of U.S. Navy Recruits on Reading Comprehension
and Selected Educationally Related Variables

Abstract

The purpose of the study was to investigate reading comprehension and related variables pertaining to the Navy's recruits during FY 81. Monthly reports provided data on high school completion, aptitude, and reading comprehension of 84,359 recruits. The data revealed that 80 percent of the recruits are high school graduates while the typical recruit reads near the eleventh grade level. Reading deficiencies were observed among 3.80 percent of the high school graduates and 2.50 percent of the non-graduates. A comparison of recruits who scored above and below the sixth grade on reading comprehension revealed distinctive differences in aptitude with the former scoring higher. Identifiable variability was also observed when the characteristics of recruits were analyzed by month across the twelve-month period.

CHARACTERISTICS OF U.S. NAVY RECRUITS ON READING COMPREHENSION AND SELECTED EDUCATIONALLY RELATED VARIABLES

Basic skills training in the U.S. Navy has become increasingly important during recent years as the characteristics of the recruit population have changed and the job demands faced by enlisted personnel have expanded. Programs have been developed for academically deficient recruits as well as personnel who are assigned to specialized skills training. Reading skill development is a component of all of these programs.

The Academic Remedial Training (ART) Program is designed to assist new recruits in acquiring the skills required to complete basic training successfully. The initial component of ART focused on reading skills needed in recruit training. Additional curricula developed for verbal skills and mathematical skills are presently undergoing field tests at the Recruit Training Center in Orlando, Florida.

Although the ART Program became operational at the Navy's three recruit training centers in 1967, standardization of the curriculum and testing procedures used in the program did not occur until 1978. A significant provision of standardization was the requirement that a commercial reading achievement test would be administered to all recruits upon entry into the Navy. The standardized program also required the use of a Computer-Managed Instruction (CMI) system that serves several training programs to perform test scoring functions and to maintain records on test performance by recruits.

The purpose of this study was to investigate reading comprehension and related variables pertaining to the Navy's recruits during the twelve-month period from October, 1980, through September, 1981. The primary focus of the study was on the classification of recruits based on their performance above or below the 6.0 reading grade level (RGL) in comprehension, the criterion used in screening recruits for ART. Other variables of interest were month of entry into the Navy, years of education, high school graduation status, and aptitude test score.

The significance of the study is two-fold. First, the data and analyses provide information on the variability of the recruit population's composition over a period of time. This information is needed by ART management personnel for efficient allocation of resources (e.g. staff, facilities, and instructional aids) in response to observable trends and/or cycles. Second, the study demonstrates potential uses of CMI system data for management functions.

Procedures

The data utilized in the study were obtained from the monthly CMI reports on population characteristics and test performance of 84,359 recruits who entered the Navy during Fiscal Year 81. Data presented in the reports include high school graduation status as well as scores from the Gates-MacGinitie Reading Comprehension Test, Level D, and the Armed Services Vocational Aptitude Battery (ASVAB). (Two series of the ASVAB - 5-7 and 8-10 - were used with the recruits who were included in the reports.)

In the reports, high school graduation status is used to compile two-factor frequency distributions of recruits who score above and below 6.0 RGL. Means are reported separately for years of education completed by the two RGL groups and by the two ASVAB series. Frequency distributions and means are presented for the RGL scores obtained from administration of the Gates to all recruits. The reports also contain means for the subtests and the Armed Forces

Qualification Test (AFQT) scores derived from the ASVAB. The AFQT score is a percentile score that is based on the word knowledge, numerical operations, and space perception subtests in ASVAB Series 5-7 and the word knowledge, paragraph comprehension, arithmetic reasoning, and numerical operations subtests in ASVAB Series 8-10.

The frequency distributions and means contained in the monthly reports were compiled for analyses. Medians were computed for the RGL scores associated with each month of the one-year period. Percentages and proportions were calculated for selected data sets. The Chi-square technique was utilized to compare various frequency distributions involving RGL scores and high school graduation status by month of entry into the Navy. Confidence interval limits were computed to compare proportions of high school graduates and non-high school graduates who scored below 6.0 RGL. The .05 level of significance was used with all inferential analyses.

Results¹

The means on years of education were compiled for recruits classified by ASVAB series used in testing and by month of entry into the Navy. Among recruits tested with the ASVAB Series 5-7, the means varied from 11.7 years during the first three months to 11.9 years during the last three months of the twelve-month period. Among recruits with whom the ASVAB Series 8-10 was used, the means ranged from 11.5 years to 11.8 years. The data revealed a high degree of consistency in means for years of education across the one-year period and between the two ASVAB series.

The means on the AFQT percentile scores were obtained for recruits classified by 6.0 RGL criterion, ASVAB series, and month of entry into the Navy. Among the recruits who had RGL scores below 6.0, the means for the AFQT ranged from a low of 31.7 (August, 1981) to a high of 36.8 (October, 1980) on ASVAB Series 5-7 and a low of 31.8 (July, 1981) to a high of 40.6 (November, 1980) on ASVAB Series 8-10. The AFQT means for recruits who had RGL scores above 6.0 varied from a low of 56.1 (June, 1981) to a high of 60.9 (January, 1981) on ASVAB Series 5-7 and a low of 55.5 (July, 1981) to a high of 60.4 (October, 1980 and January, 1981) on ASVAB Series 8-10. The differences in the means for the high and low RGL groups were generally 20-25 percentile points when compared by month or ASVAB series. These differences suggest that the two RGL groups are relatively distinct with respect to AFQT scores.

The medians and means on RGL for recruits by month of entry into the Navy were determined. The medians showed a slightly increasing trend from a low of 10.7 RGL for the first month to 11.3 RGL for the eleventh month of the one-year period. The means were relatively stable with a range from 9.7 RGL for the first month to 10.1 RGL for the eighth month. The relative magnitudes of the medians and the means indicated that the score distributions were negatively skewed. This phenomenon is an artifact of the population composition and the RGL range for the test norms (2.4 RGL to 12.5 RGL).

The frequency and percentage distributions of recruits by high school graduation status and month of entry into the Navy were obtained. The data for the total group indicated that approximately 78 percent of the recruits were high school graduates. The percentages of recruits who were graduates varied considerably across the twelve-month period with the range being from

71.57 percent in May, 1981, to 81.66 percent in October, 1980. After an initial decrease in percentages occurred from October to November, 1980, the percentages were relatively stable through April, 1981, except for January, 1981 which had an elevated percentage. Following a decrease in May, 1981, the percentages increased for the remaining four months. The Chi-square comparison of the distributions on high school graduation status by month revealed that the distributions differed very significantly (Chi-square = 433.131, $p < .005$).

The frequency and percentage distributions were prepared for high school graduate recruits by 6.0 RGL criterion and month of entry into the Navy. Within the composite group, 3.8 percent of the high school graduates scored below 6.0 RGL on comprehension. The percentages of high school graduate recruits who scored below 6.0 RGL ranged from 2.92 percent in January, 1981, to 5.35 percent in April, 1981. A general downward trend in percentages can be observed during the first four months of the period with a reversal occurring over the next three months followed by a pattern of lower percentages. The Chi-square analysis of distributions of high school graduate recruits classified by 6.0 RGL criterion and month of entry indicated that statistically significant differences existed (Chi-square = 71.110, $p < .005$).

The frequency and percentage distributions of non-high school graduate recruits were compiled by 6.0 RGL criterion and month of entry into the Navy. The data for the composite group showed that 2.50 percent of the recruits who were not high school graduates obtained scores below 6.0 RGL on comprehension. Across the twelve-month period, the percentages of non-high school graduate recruits who scored below 6.0 RGL varied from a low of 1.51 percent in February, 1981, to a high of 3.79 in June, 1981. A slightly declining trend in percentages occurred during the five months beginning in October, 1980, followed by a general pattern of increase. The results of the Chi-square comparison of distributions of non-high school graduate recruits classified by 6.0 RGL criterion and month of entry showed significant differences (Chi-square = 35.733, $p < .005$).

The frequency and percentage distributions were tabulated for recruits who scored below 6.0 RGL classified by high school graduation status and month of entry into the Navy. The data for the twelve-month period indicated that 84.65 percent of the recruits who scored below 6.0 RGL were high school graduates. The monthly data showed that the percentages of high school graduates among the recruits scoring below 6.0 RGL had a range from 76.04 percent in June, 1981, to 89.32 percent in October, 1980. A trend of relatively stable percentages occurred during the first seven months of the period followed by decreases the next two months and relative stability at the earlier level during the last three months. The Chi-square comparison of the distributions of recruits scoring below 6.0 RGL classified by high school graduation status and month of entry revealed significant differences (Chi-square = 32.756, $p < .005$).

Comparative analyses were performed on the proportions of recruits who scored below 6.0 RGL classified by high school graduation status and month of entry into the Navy. The data on the proportions of recruits scoring below 6.0 RGL are discussed above for high school graduates and non-high school graduates separately. The relative proportions of the two groups showed that a higher proportion of the high school graduates than non-high school graduates scored below the 6.0 RGL criterion during 11 of the 12 months (except for June,

1981) as well as for the twelve-month period. The proportions of the two groups who had RGL scores below 6.0 were compared by computing confidence limits for the differences in proportions using the .95 level of confidence. The results indicated that the overall proportions differed significantly since the confidence limits did not include the 0 within the range. The proportions also differed significantly for six of the twelve months. In all cases where significant differences were found, the proportions of high school graduates who scored below 6.0 RGL were greater than the corresponding proportions for non-high school graduates.

Conclusions

The data presented in this study provide selected descriptive information about the Navy's recruit population for a twelve-month period of time. The results obtained by compiling and analyzing the data constitute the basis for the conclusions related to the recruit population.

The educational background of the recruits indicates that four-fifths are high school graduates. Since the data on years of education represent the years completed, the monthly means place the group as a whole in the middle to higher twelfth grade range. Higher percentages of the recruits who enter the Navy in the summer and early fall months are high school graduates than in other months of the year. This increase might be expected in the months following the traditional high school graduation period of the year.

The aptitude data on recruits as reflected by the AFQT scores do not show any clearly discernible seasonal or cyclic pattern. The means indicate, however, that recruits who perform above 6.0 RGL are likely to score considerably higher than recruits who obtain scores below 6.0 RGL. The monthly means are slightly more variable for the lower reading group than for the higher reading group.

The performance of recruits on reading comprehension shows a gradual improvement over the twelve-month period of this study. This trend is evident from the monthly medians, means, and percentage distributions of the RGL scores. Based on measures of central tendency, the monthly medians reveal an increase from the tenth grade to the eleventh grade while the means are approximately one grade lower.

Among the recruits who score below 6.0 RGL, six-sevenths (85 percent) are high school graduates as compared to four-fifths (78 percent) of the entire recruit population being high school graduates. The data reveal that 1.30 percent more of the high school graduates than non-high school graduates score below 6.0 RGL. This difference is significant for the two composite groups and relatively consistent across the twelve-month period. The percentages of each group tend to decline through the first half of the year and show some increase subsequently.

The conclusions suggest that variability in the recruit population can be expected during a twelve-month cycle. Factors that may reveal differences in distributions across the twelve-month period are RGL scores, high school graduation status, high school graduates and non-high school graduates who score above and below the 6.0 RGL criterion, and high school graduation status of recruits who score below 6.0 RGL. This information can be useful to AR program management personnel in planning the allocation of resources to serve the students assigned to the program in the most efficient manner possible.

Characteristics of U.S. Navy Recruits

The data utilized in the study are available to management personnel in the monthly CMI reports on Navy recruits. By constant monitoring of the data presented in these reports, ART management personnel can anticipate more accurately the fluctuations that will occur in the flow of students to the ART Program.

Footnote

¹Tables presenting the data and analyses discussed are available from the authors.

VALIDATION OF THE COMPUTER PROGRAMMER APTITUDE BATTERY

J.P. McMenemy

Canadian Forces Personnel
Applied Research Unit

The Canadian Forces adopted the Computer Programmer Aptitude Battery (CPAB) (Palormo, 1967, 1974) as an aid in the selection of Computer Programmer Trainees in 1973. The selection model was developed by the Canadian Forces Personnel Applied Research Unit (CFPARU) and was based on research carried out by Martin (1971, 1973). Since 1973, the ADP environment in the Canadian Forces has changed significantly. The trainee population has expanded to include NCOs as well as officers and a second programmer training course has been developed. This paper reports a study of the predictive validity of the CPAB against the newest Canadian Forces ADP training programme, the Maritime Tactical Programmers Course (MARTAC) at the Canadian Forces Fleet School, in Halifax, Nova Scotia. This represents part of a general reassessment of programmer selection.

The course lasts 45 training days and is broken down into three main areas: coding (80%), programme design (10%), and problem solving (10%). Students learn two command packages, CMS and ULTRA-16. The MARTAC course concentrates upon computer system design and application so that graduates can use ADP equipment and programmes to solve tactical problems, mainly in anti-submarine warfare.

METHOD

Subjects

The subjects were officer and NCO trainees on the Maritime Tactical Programmer Course between 1978 and 1981. All were from Naval or Air Force operational or technical backgrounds. Criterion data were available on 114 trainees but test scores were available on only 83. The bulk of the 31 missing test scores are unavailable either because test scores were not forwarded to CFPARU, or, which occurred more often, the test was not administered. About one-third of the trainees are commissioned officers of the rank of Major and below, and the rest are NCOs, from the rank of Master Corporal or above. Master Corporal is the first level of supervisor among the other ranks.

It was not possible to identify officers and other ranks for this analysis. About half the officer graduates are posted to programming jobs within Maritime or Air Command. The remainder go to operational or technical support positions where ADP training is an asset. Very few of the other ranks become programmers. Most are employed in technical support positions where a knowledge of computer systems and ADP is either necessary or very desirable.

The Test

The CPAB, published by SRA, (Palormo, 1967, 1974) is composed of five multiple choice subtests: Verbal Meaning, Reasoning, Letter Series, Number Ability, and Diagramming. A score consisting of the number of items answered correctly is computed for each subtest and the total battery. The test was administered by Base Personnel Selection Officers and results were forwarded to CFPARU for keypunching and storage on magnetic tape.

At some point instructors were permitted access to test scores. Instructors claim, however, that they seldom examine the test scores. The course Director states that on rare occasions scores will be examined after the third week of the course as a diagnostic and counselling aid if a student is having difficulty with the material. Regardless of instructors' disclaimers, this might be classified as a possible contaminant in the study and will be discussed in a later section of the paper. No test cut-off was specified for selection to the training programme¹.

The Criterion

Students are evaluated by three tests and a two part problem. The tests are administered at the end of each of the first three weeks and constitute 20% of the overall evaluation. The remainder of the assessment is a comprehensive problem which students take two to three weeks to complete. Part I of the problem is worth 40% and Part II, 20% of the final grade. Course reports summarize the instructors' assessment as follows:

- A - competent to be employed as an independent programmer;
- B - gained sufficient knowledge and skill to be employed as a programmer but should be provided 6-12 months of supervision before becoming an independent programmer;
- C - gained some understanding of the material but is not recommended for employment as a programmer; and
- F - failed to meet minimum requirements.

Analysis

The relationship between CPAB scores and course grade was investigated primarily by calculating the product moment correlation between the test scores and course grades. The grading scheme (A,B,C,F) was transformed into a four point scale, with 4 being the highest grade (A) and 1 representing a failure (F). Simultaneous multiple regression was then applied to generate the regression equation for later cross-validation. It is recognized, however, that multiple regression results must be interpreted cautiously because the number of subjects is relatively small.

RESULTS

The means and standard deviations of CPAB scores show little evidence of restriction of range. The trainees' mean on the CPAB total is 86.08 and the standard deviation is 23.26. Total score ranged from 35 to 132. Means and SDs for the subtests and CPAB total are presented in Table 1. Local reliability estimates could not be computed since only subtest and CPAB total scores are available.

¹ Personnel Selection Directive-18/75 states: "Two candidates who achieved raw scores of 76 and 74 have gone on to become competent programmer/analysts. To this end the preferred minimum raw score on the CPAB is 75, however no cut-off shall be applied by PSOs."

Table 1: Means and Standard Deviations for the CPAB				
	Mean		SD	
Verbal Meaning	20.10	(17.3)*	6.81	(6.8)
Reasoning	11.57	(10.9)	5.76	(5.9)
Letter Series	13.59	(13.4)	4.16	(4.7)
Number Ability	14.63	(13.1)	4.87	(5.3)
Diagramming	25.78	(23.8)	7.42	(8.8)
CPAB Total	86.08	(78.4)	23.26	(25.2)
N = 83				
* bracket indicates publishers norms for trainees - N=641 - source CPAB Examiners Manual Second Edition, 1974.				

The sub-test intercorrelations (Table 3) reveal that the trainees subtest scores are highly correlated. In several cases, the correlations are markedly higher than the publisher's estimates or those found by Martin (1973).

Table 2: CPAB Intercorrelations and Reliability Estimates						
	1	2	3	4	5	6
1 Verbal Meaning	(.86) ^a	.57 ^b	.48	.53	.33	
2 Reasoning	.70 ^c	(.88)	.60	.67	.53	
3 Letter Series	.36	.54	(.67)	.56	.48	
4 Numberability	.57	.71	.52	(.85)	.38	
5 Diagramming	.48	.66	.57	.49	(.94)	
6 CPAB Total	.79	.89	.70	.79	.83	(.95)
^a diagonal - publishers reliability estimates						
^b upper half - publishers intercorrelations-source CPAB Manual						
^c lower half - MARTAC trainees' intercorrelations-source CPAB Manual						

The course grades of those subjects for whom both test scores and grades are available were examined. No test results were available for any of the course failures although it is known that four F grades were recorded. The grade range therefore is 2 to 4 with a mean of 3.13 and a standard deviation of 0.84.

Product moment correlations between CPAB subtest and total scores and course grade were all positive and significant ($p < .01$). It can be noted in Table 4 that the highest correlation was between the Reasoning subtest and grade ($r = .5688$). CPAB total correlated almost as high at .5588.

Table 3: Validity Coefficients - CPAB and Course Grade			
Verbal Meaning	.4128	Number Ability	.4269
Reasoning	.5688	Diagramming	.5079
Letter Series	.3194	CPAB Total	.5588

The subtest scores were entered into a simultaneous multiple regression analysis and yielded a multiple R of .5992. Simple correlation of total score with course grade indicates that the test accounts for 31.2% of criterion variance. The R^2 yielded by multiple regression was .359. Apparently, since the subtests are highly correlated, relatively little precision is gained by considering the subtests individually rather than using total score.

Table 4: Multiple Regression - CPAB Test Scores and Course Grade			
	unstandardized B. weight	standardized beta	contribution
Verbal Meaning	.009	.00734	.17037
Reasoning	.057	.39020	.15347
Letter Series	-.015	-.07308	.00023
Number Ability	.010	.05869	.00102
Diagramming	.030	.26054	.03393
Constant 1.72			
	R^2 .35902	R .59919	

The relationships found in this study are considerably higher than previously found in Canadian Forces studies of the CPAB. The possible explanations for this relate directly to the interpretation and application of results of this study.

DISCUSSION

The validity coefficient found in this study are impressive but must be interpreted with some caution for several reasons. The small number of subjects may not provide a stable estimate of validity. Secondly, instructors access to CPAB results may have artificially inflated the correlation between the CPAB test performance and course grade.

The last point is the most troublesome and will be considered first. It is possible that providing instructors access to CPAB results established an expectancy which affected the instructor-student relationship during training and/or the assessment process. The instructors claim this is not so but the effect is likely to be so subtle that instructors would not be aware of it, even if it were occurring. If the instructors may be taken at their word, that is, results were only consulted on rare occasions when students were obviously having problems, it can be argued that an expectancy was established by student performance. The self-fulfilling prophecy effect would then be negligible and a greater degree of trust could be placed in the results of the study.

The likelihood that a near zero relationship could be inflated to the degree shown in this study is very low, especially since the test's validity has been been verified in similar settings (Martin, 1971, 1973). The possibility of an effect cannot be discounted, but neither can its magnitude be estimated. For a definitive answer, one must await a cross-validation study.

The coefficients of determination generated by simple correlation and multiple regression, .31 and .36 respectively are markedly greater than previously seen ($r^2=.13$, $r^2=.09$) (Martin 1971, 1973) in the Canadian Forces. This can be attributed to inflation by instructors knowledge of test scores but can also be accounted for by other factors. Validity coefficients calculated by Martin (1971, 1973) were probably suppressed by restrictions of range. The 1971 study was a concurrent validity study so that all subjects were pre-selected and had experience at some of the types of problems presented in the test. This study examined military personnel and civilian employees of the Department of National Defence and it was found that the test was much more predictive of military members performance than of civilians. One explanation is that the military people were rated more reliably than the civilian employees, but it is also possible that they constituted a more heterogeneous group. The civilian ADP workers would apply for and be hired

for the programmer jobs only if they had some interest in, and presumably, some talent for the job. The military people that are employed in the environment are selected, in part at least, because of their experience and success in the military. This factor would tend to expand the range of interest in, and aptitude for programming. Presumably, a similar effect would be found among MARTAC candidates. Since trainees were not selected on the basis of the CPAB, the range of predictor variance could be expected to be greater than that usually found among ADP trainees and employees. Greater variability in the predictor variable would, given some degree of validity, yield higher correlations than the same predictor with the lower end of the range eliminated.

The small number (83) available for this study reduces the confidence one may have in the stability of results. The correlations however, are very strong and indicate that the test is basically valid.

CONCLUSION

The CPAB is a valid test for use in the selection of MARTAC trainees. The application of the test for selection of the trainees has not yet been decided. This study is part of a larger assessment of ADP selection procedures and will be applied only in the larger context. Other factors to be considered include the effects of military rank, education, previous military training, and occupational background. As well, consultation must be carried out with career management staff and higher levels within Maritime Command.

The use of tests like the CPAB can be justified only to the extent that they contribute to more valid personnel decisions. This can only be accomplished in the context of organizational needs and values. To this end, information relating to costs, and benefits of possible training outcomes will be assembled and analyzed. Then an appropriate application of decision theory (Cronbach and Gleser, 1965) can be devised.

REFERENCES

1. Cronbach, Lee J. and Gleser, G.C. Psychological Tests and Personnel Decisions. Urbana: University of Illinois Press, 1965.
2. Martin, M.A. A study of the Concurrent Validity of the Computer Programmer Aptitude Battery. Studies in Personnel Psychology, 1971, 3, 69-76.
3. Martin, M.A. Validity of the Computer Programmer Aptitude Battery - Canadian Forces 1973. Technical Report 73-7, Canadian Forces Personnel Applied Research Unit, Willowdale, Ontario, 1973.
4. Palormo, Jean M. Computer Programmer Aptitude Battery Manual (1st Edition) Chicago, Science Research Associates, 1967.
5. Palormo, Jean M. Computer Programmer Aptitude Battery Manual (2nd Edition) Chicago, Science Research Associates, 1974.
6. Personnel Selection Directive 19/75. National Defence Headquarters (5763-2 DPAR) 01 September 1975.

ABSTRACT

Validation of the Computer Programmer Aptitude Battery

J.P. McMenemy

Canadian Forces Personnel
Applied Research Unit

The Computer Programmer Aptitude Battery was administered to 83 Canadian Military Officers and NCOs prior to the Maritime Tactical Programmer Course. Correlation and Multiple Regression demonstrated the validity of the test. In Multiple Regression Analysis the test accounted for 35.9% of criteria variance. A possible confound in the study and the development of a comprehensive selection model are briefly discussed.

VIBRATION EFFECTS ON DIGIT-SYMBOL CODING*

Alvah C. Bittner, Jr., John C. Guignard
Jeffrey C. Woldstad, and Robert C. Carter

Naval Biodynamics Laboratory, New Orleans, LA. 70189

Abstract

A Digit Symbol Coding task was investigated with three trials each before, during, and after whole-body sinusoidal vibration. Four inexperienced and two experienced subjects were tested at 8 Hz at 0.21 g_{rms}. Results indicated that individual differences and related (group)² interactions accounted for respectively 51.1% and 5.0% of the total sum-of-squares. Simple vibration effects accounted for 14.8% of the variance. The present and previous results indicate: (1) performance decrement reductions with practice under vibration, and (2) the operation of an experience-practice interaction on performance after vibration. It was concluded that vibration research should utilize control groups, randomization, and repeated measures methodologies. With regard to the international standard on human whole-body vibration exposure, ISO 2631-1978, it was recommended that research be directed at the under-explored realm of psychophysiological and biomechanical effects during long-term exposures.

An experimental program is underway at the Naval Biodynamics Laboratory to measure whole-body vibration effects in man. These studies are aimed at establishing correlations between the psychophysiological responses and the biodynamic (inertial) responses of human volunteer subjects to various mechanical input forces of the kind to be experienced by the crew of ships and aircraft. As part of this program, a series of investigations is being conducted using vertical (z-axis) sinusoidal whole-body vibration. This report is concerned with the effects of (8 Hz, 0.21 g_{rms}) vibration on performance of a Digit-Symbol Coding Task.

Digit-Symbol Coding has a long history of investigation and development. Initially developed by Otis around World War I, the task is a part of several intelligence batteries (Wechsler, 1958). Coding has been recently reviewed, examined for stability, shown sensitive to ship motion environments, and recommended for inclusion in an environmental assessment battery (Pepper, Kennedy, Bittner, & Wiker, 1980). Carter and Sbisa (1981) have more recently developed a computer based method for generating great numbers of alternate forms of (Digit-Letter) Coding.

Guignard, Bittner, and Carter (1981) have studied Digit-Letter Coding under various vibration conditions. In their investigation, they noted an interaction between vibration experience and performance following vibration exposure. Specifically, a group of inexperienced subjects performed at near experienced group levels before and during exposure, but fell off substantially afterwards ($F(4, 32) = 3.82$; $p < .02$). Guignard et al. also studied "Spoke" (Bittner, Lundy, Kennedy, & Harbeson, in press) and "Aiming" (Fleishman & Ellison, 1962) manual

*This study was conducted under U.S. Navy Bureau of Medicine and Surgery Work Unit MOC96-PN002-1018. The opinions are those of the authors and do not necessarily reflect those of the Department of the Navy.

task but found no similar interaction for either manual task (p > .41). The pattern of results suggested carry-over effects on mental (Coding) but not on manual tasks.

The primary goal of the present study was to research vibration effects on Digit-Symbol Coding. A secondary goal was to evaluate a basic paradigm for investigating within exposure conditions (Before, During, and After). The purpose of this investigation was to obtain further evidence pertaining to the validity of the current international standard, ISO 2631-1978, on human exposure to vibration (International Standards Organization, 1978).

Method

Subjects and Groups

The subjects were 6 Navy enlisted men (aged 19 to 24) who had volunteered for duty as biodynamics research subjects. They had been selected to be unusually free of skeletal, cardiopulmonary and other medical or psychological conditions which would preclude participation in potentially hazardous environmental research. The subjects were otherwise typical of the general enlisted population. All subjects were recruited, evaluated, and employed in accordance with SECNAV Instruction Series 3900.39 and BUMED Instruction Series 3900.6. These instructions are based upon informed voluntary consent and meet provisions of prevailing national and international guidelines regarding proper human experimentation. A more detailed description of the volunteers and their selection is given by Thomas, Majewski, Ewing, and Gilbert (1978).

The subjects were divided into two experience-level groups. The first group (1) consisted of 4 subjects who had had no previous formal experience with performance tests under vibration. The second group (2) consisted of two subjects who had had previous exposure to experimental vibration and experience with the task and conditions of this study.

Vibration Conditions and Procedures

The sinusoidal vertical vibration condition employed in this study was 8 Hz at 0.21 g rms delivered at the seat and foot-rest for 8 minutes. This condition was² selected with reference to ISO 2631-1978 and the level applied to the seat was twice the value of the "Fatigue/Decreased Proficiency Boundary" defined for short-term, whole-body exposure. The exposure condition was one previously evidencing simple vibration and carry-over effects (Guignard et al., 1981).

Prior to vibration exposure, subjects were given 5 trials of Digit-Letter Coding in a training session on the day preceding their experimental run. Equivalent forms of the task were employed which had been generated by the program of Carter and Sbisa (1981). On the experimental day and before boarding the vibration machine, subjects performed the tasks twice to "warm up". Once mounted on the vibration machine, three data trials were administered each Before (B), During (D), and After (A) vibration for a total of 9 observations. The duration of each trial was 120 seconds and the score was the number correct. Between trials, subjects had 40 seconds of rest. Trials were arranged so that the first "After trial" was initiated simultaneously with the cessation of vibration. The training session and warmup procedures were similar to those employed by Guignard et al. (1981).

Subjects were debriefed subsequent to experimental data collection. As part of this debrief, it was noted that Coding Task could be divided

into three subtasks: "Seeing", "Thinking", and "Writing". Subjects were asked to directly estimate the "proportion of capacity which remained under vibration" for each subtask independently. A modulus of 50 was designated "as the capacity under no vibration". Subjects were given scaling illustrations of 25 for a reduction to half capacity and 75, to represent a 50% increase in capacity. The subjects' scaling comprehensions were checked prior to their estimations. The direct estimation procedure was derived from Stevens (1975).

Apparatus

The subjects rode seated on the Laboratory's 28,000 lbf electrodynamic vibration machine, operating in its vertical (z) mode (i.e., vibrating in the direction of gravity), and equipped with a rigid seat and foot-rest directly coupled to the armature of the machine. The hard seat was, for comfort, shaped similar to a farm tractor seat. It also incorporated the seat reference accelerometer used to monitor the vibration input to the subject in his z-axis. Only gravity was used for restraint: there were no straps or back-rest. The machine is capable of shaking a seat and human subject without extraneous mechanical support or appreciable distortion of the vibration wave-form in the conditions studied: rms acceleration at the seat was controllable to within $\pm 1\%$; vibration frequency was controllable to within $\pm 0.5\%$; and total harmonic distortion was negligible. Subjects were provided with clip-boards for holding task materials and with fine-point marker pens to enter responses. Tasks were presented and timed by a test administrator who stood beside the vibration machine.

Results

The analysis was conducted in two phases. In the first phase, a repeated measures analysis of-variance (ANOVA) was conducted on scores from Digit-Symbol Coding. In the second phase, the debrief scaling estimates was analyzed by ANOVA. These two phases are described below.

Coding Analysis

Table 1 summarizes the ANOVA. In this analysis, a Group factor (G) divides the subjects into inexperienced (Group 1) and experienced (Group 2). The 9 repeated measures on each subject represents 3 trials (T) on each of the three vibration Conditions (C): Before (B), During (D), and After (A). Table 1 shows that each main effect and interaction contrast within-subject (Ss) was examined. In this table and following Bittner and Carter (in press), C(1) and C(2) are linear and quadratic contrasts between Conditions ((A-B) and ((B+A)/2 - D)), and T(1) and T(2) are linear and quadratic components for Trials. Experimental interest was focused on the individual contrasts.

Table 1 reveals four effects clearly significant ($p < .05$) and two of marginal significance ($p < .07$). These can be viewed in three clusters of respectively one, two, and three effects. First, Subjects- Within-Groups (Ss w.G) shows a very highly significant indication of (within-group) individual differences ($F(4,16) = 33.82$; $p < 10^{-5}$). This (Ss w. G) is also a substantial effect representing 51.1% of the total sum-of-squares (SS).

C(1) and C(1)G constitute the second cluster of effects. In the current context, C(1)G is the interaction of the (A-B) difference with Groups. C(1) and C(1)G respectively show: a marginally significant ($F(1,8) = 4.87$; $p < .06$) drop in performance across groups; and a clearly significant indication of Group differences ($F(1,8) = 5.34$; $p < .05$). Together, these effects make up 6.7% of the total SS. Figure 1 illu-

TABLE 1. Analysis-of-Variance Summary

SOURCE	SUM-OF-SQUARES	DEGREES OF FREEDOM	F	PROBABILITY
<u>Between Ss:</u>				
Group (G)	5.33	1	0.01	NS
<u>Ss w. G</u>	1561.50	4	33.82*	<.00001
<u>Within Ss:</u>				
Conditions (C)				
C(1)	98.00	1	4.87	.05 < p < .06
C(2)	450.67	1	22.39	<.002
C(1)G	107.56	1	5.34	<.05
C(2)G	16.67	1	0.83	NS
<u>Ss w. CG</u>	161.00	8		
Trials (T)				
T(1)	6.13	1	0.16	NS
T(2)	3.38	1	0.09	NS
T(1)G	6.13	1	0.16	NS
T(2)G	12.04	1	0.32	NS
<u>Ss w. TG</u>	298.83	8		
CT Interactions				
CT(1,1)	16.33	1	1.42	NS
CT(1,2)	6.25	1	0.54	NS
CT(2,1)	64.00	1	5.55	<.032
CT(2,2)	4.08	1	0.35	NS
CT(1,1)G	3.00	1	0.09	NS
CT(1,2)G	0.03	1	0.00	NS
CT(2,1)G	4.00	1	0.35	NS
CT(2,2)G	44.08	1	3.82	.05 < p < .07
<u>Ss w. CTG</u>	184.67	16		

* Conservatively Tested Against Ss w. CTG

strates both of these effects where: Group 2 scores well Before (B) but shows a moderate drop After (A) exposure; and Group 1 shows a very slight (A-B) increase. Overall, the groups show a modest (A-B) decrease.

The third cluster consists of C(2), CT(2,1), and CT(2,2)G which altogether constitute 18.3% of the SS. C(2), ((B+A)/2 - D), shows a highly significant indication of an overall effect due to vibration ($F(1,8) = 22.39$; $p < .002$). In addition, CT(2,1) indicates that the "vibration effect" changes linearly between the first and last Trials across Groups ($F(1,16) = 5.55$; $p < .032$); and CT(2,2)G provides a marginally significant indication that the Group's vibration effects differ in their quadratic component across Trials ($F(1,16) = 3.82$; $p < .07$). Figure 2 illustrates all of the effects of this cluster in terms of Base line ((B+A)/2) minus vibration (D) differences. Examining this figure, it can be seen that all differences shown in the figure are positive; vibration effects are present across all trials and Groups as indicated by the significant C(2) effect. Next, it can be seen that there is a reduction in the Baseline-Vibration difference between Trial 1 and Trial 3; this gave rise to the CT(2,1) effect. Lastly, the figure indicates that the decrease in Baseline-Vibration occurs more slowly for Group 1 than for Group 2. Overall, the results of this cluster suggest: practice reduces the vibration effect; and previous experience enhances this reduction.

Direct Scaling Analysis

Subject proportional estimates for "Seeing", "Thinking", and "Writing" were calculated and cast into a two-way repeated measures table prior to analysis. ANOVA of this table revealed no significant differences between subjects ($F(5,10) = 1.29$; $p > .34$) and highly significant task differences ($F(3,10) = 6.05$; $p < .013$). Thinking was seen to have the greatest remaining capacity (0.883), Seeing a lesser capacity (0.720), and Writing the least remaining capacity (0.567). Significantly, 5 of 6 subjects also specified Writing as their greatest problem prior to estimation instructions ($p < .02$). Subject estimation indicated that vibration effects were largely input-output in nature (Seeing-Writing) with the greatest problem being output (Writing).

Discussion

This investigation was directed at effects of vibration on Digit-Symbol Coding. A subsidiary concern was with the paradigm employed in this study. The purpose of this investigation was to gain further insights into the current international standard (ISO 2631-1978). An overview of the findings and implications of this investigation will be discussed.

Overview

Individual-differences and vibration effects were the most salient findings of this study. Overall individual differences, as accessed by Subjects-within-Groups (S_s w. G), constituted the majority of the total sums-of-squares (51.1%). This proportion, it is noteworthy, is comparable with the 29.9% reported for overall subject differences by Guignard, Bittner, and Carter (1981). In addition, Group differences were manifested in interactions with conditions (C(1)G) and trials (CT(2,2)G) with additionally explained sums-of-squares (5.0%). Main effects of vibration (C(2)) accounted for 14.8% of the total sum-of-squares. The direct estimation results indicated that the vibration (C(2)) effect was largely due to "Writing"; this supported the correlational analysis of Guignard et al. (1981) who found the effect on Coding was essentially mechanical, not mental in nature. Altogether, individual-differences related sums-of-

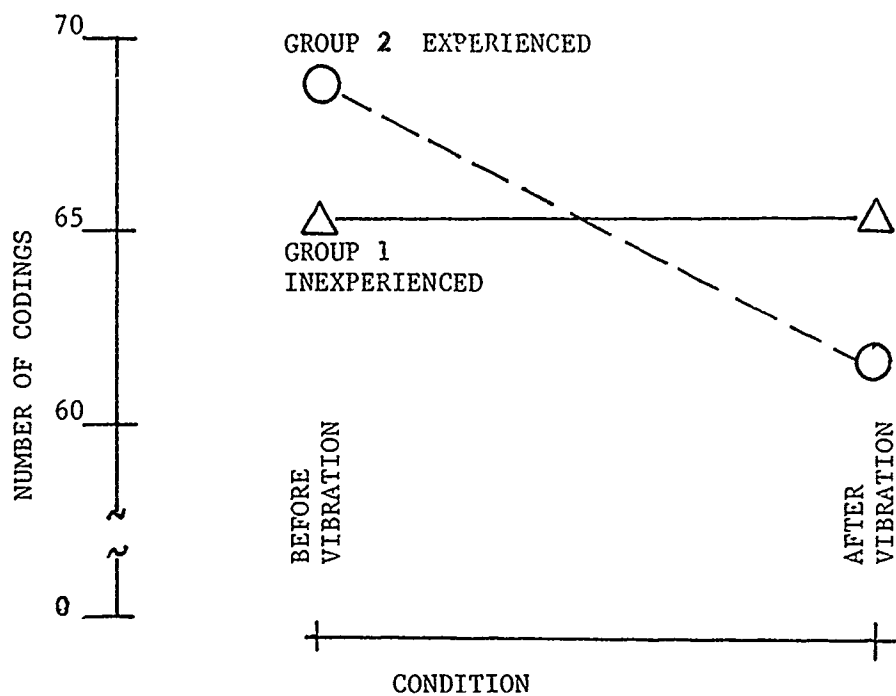


Figure 1. Group Scores Before and After Vibration

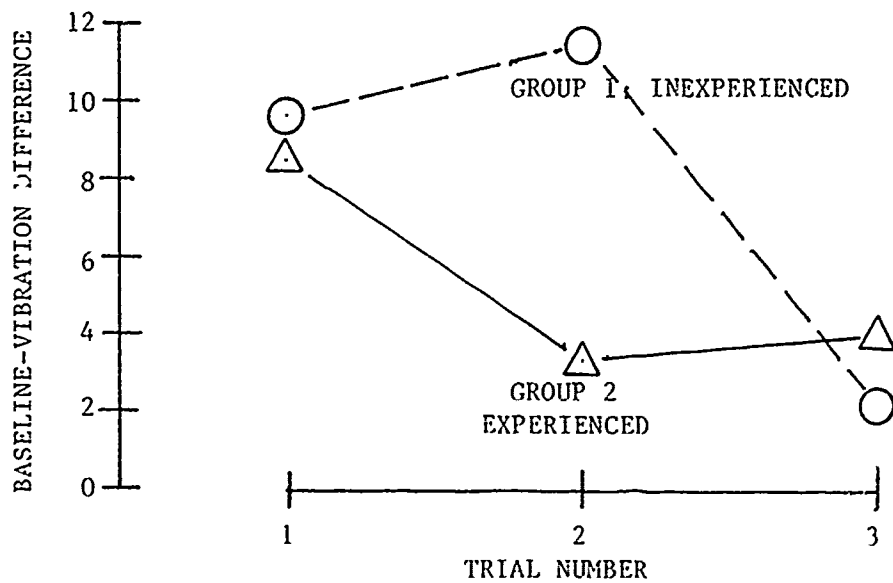


Figure 2. Group Baseline-Vibration Differences Over Trials

squares accounted for 56.1% and simple vibration effects for 14.8% of the total. Individual-difference related sources clearly dominated the explained variation.

Figures 1 and 2 illustrate the nature of the group interactions. Figure 1 shows that, while the inexperienced subjects (Group 1) showed improved performance subsequent to vibration, experienced subjects (Group 2) showed a decline. This finding appears, at first glance, to be in conflict with the report of Guignard et al. (1981) who found that their inexperienced group showed a decline (Before-After) while their experienced group demonstrated improvement. However, reexamination of Guignard et al.'s. (Table 5) data revealed increases early in practice between measures Before and After (e.g., Exp. 1), but that declines in later experiments (Exp. 4). From current and previous findings, a complex interaction of experience and practice might be posited; however, additional empirical evidence appears preferable to speculative attempts at explanation using the current data base. Figure 2 shows both groups improved under vibration performance with practice; however, inexperienced subjects were slower to improve than experienced. Interestingly Guignard et al. (1981) were unable directly to demonstrate improvements under vibration as they only had single trials under each condition. However, comparison of their experiments 1 and 4 (8 Hz, 0.21 g_{rms}) results reveal an apparent reduction of the overall Baseline-Vibration difference from 15.5 to 12.0 for repeated testings with several months separation. Thus, the results of Guignard et al. (1981) are not inconsistent with those of the present experiment and indicate task learning under vibration. Altogether, the present and previous results indicate: reduction in vibration performance decrements with practice and an interaction of experience and practice with performance after vibration.

Implications.

The individual-difference related results have implications for further investigations. Following Guignard et al. (1981), the consistent large scale differences between subjects across conditions suggest the continued use of repeated measures designs for their control. In addition, the group related interactions in this study also support their recommendation that subjects be given a modicum of vibration exposure before formal experimentation. However based upon the current investigation, justification for this procedure would involve minimization of possible effects of experience-practice interactions. The need to control individual-difference related interactions reemphasizes the (Fisher, 1966) requirement for random assignment of subjects to experimental conditions. In addition, individual-difference related interactions suggests the addition of a control group, particularly where interactions with the exposure conditions are involved. Campbell and Stanley (1963) have described the value of control groups and the pitfalls of research without them. Altogether, the individual-difference related results emphasize the need for controls, randomization, and repeated measures.

Trial related effects highlight the need for study of the development of effects during vibration. The performance adaptation to vibration which has been shown suggests the operation of many possible mechanisms: psychological, physiological, and biomechanical. Such adaptations could mean that the international standard (ISO 2631-1978) is unduly conservative for long-term exposures. Based upon assessment of the meager data base for extended exposures, this possibility has been previously

pointed out by Guignard et al. (1981). The need to examine the development of effects and the paucity of data at longer exposures recommends their joint consideration in future research. Fortunately, time-series and other methodologies applicable to the study of dynamic long-term changes are currently available (cf, Bittner & Carter, in press).

It is concluded that future research should utilize control groups, randomization, and repeated measures methodologies. With regard to ISO 2631-1978, it is recommended that future research be directed at the under-explored realm of long-term psychological, physiological, and biomechanical changes.

References

- A.C. Bittner, Jr. and R.C. Carter. Repeated measures of human performance: A bag of research tools. In: J.C. Guignard (Ed.). Proceedings of the International Workshop on Research Methods in Human Motion and Vibration Studies, in press.
- A.C. Bittner, Jr., N.C. Lundy, R.S. Kennedy and M.M. Harbeson. Performance Evaluation Tests for Environmental Research (PETER): Spoke Tasks. Perceptual and Motor Skills, 1982, in press.
- D.T. Campbell and J.C. Stanley. Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally, 1963.
- R.C. Carter and H. Sbisà. Human Performance tests for repeated measurements: Alternate forms of eight tests by computer. New Orleans: Naval Biodynamics Laboratory, Unpublished manuscript, 1981.
- R.A. Fisher. Design of Experiments (8th Ed). New York: Hafner, 1966.
- E.A. Fleishman and G.D. Ellison. A factor analysis of fine manipulative tests. Journal of Applied Psychology, 1962, 46, 95-105.
- J.C. Guignard, A.C. Bittner, Jr., and R.C. Carter. Methodological investigation of vibration effects on performance of three tasks. Proceedings of the 25th Annual Meeting of the Human Factors Society, Rochester, N.Y., October 1981, 342-346.
- International Organization for Standardization (ISO). Guide for the evaluation of human exposure to whole-body vibration. Geneva: ISO, 1978. (International Standard ISO 2631-1978).
- R.L. Pepper, R.S. Kennedy, A.C. Bittner, Jr., and S.F. Wiker. Performance Tests or Environmental Research (PETER): Code Substitution Test. Proceedings of the Seventh Psychology in the DOD Symposium. Colorado Springs: USAF Academy, 1980, 451-457.
- S.S. Stevens. Psychophysics: Introduction to Its Perceptual, Neural and Social Prospects. New York: Wiley, 1975.
- D.J. Thomas, P.L. Majewski, C.L. Ewing, and N.S. Gilbert. Medical qualification procedures for hazardous duty aeromedical research. AGARD Conference Proceedings No. 231. Neuilly-Sur-Seine, France: AGARD, 1978, A-3: 1-13.
- D. Wechsler. The Measurement and Appraisal of Adult Intelligence. Baltimore: The Williams & Wilkins Co., 1958.

The Celestial Training Device - An Example of a Low Fidelity, Highly Effective Part Task Trainer

INTRODUCTION

Over the past decade, considerable attention has been given to the development and use of training simulators. Simulator technology has allowed great advances in the ability to realistically simulate the operational environment. These advances have been manifested in all areas of simulation technology from visual displays (e.g., Yan, 1980) to maintenance trainers (Smith, 1980).

Currently, as these technological advances have been occurring, behavioral research has indicated that a high level of physical fidelity is not required for training many types of tasks. For procedural tasks in particular, a number of studies have indicated that non-operational mock-ups or photographs frequently allow as much transfer of training to the operational equipment as a dynamic training device (e.g., Proplet and Boyd, 1970; Sitterly and Berge, 1972; Cox, Wood, Boreu and Thorne, 1965; Johnson, 1978). Even for psychomotor skills, there is some doubt regarding the fidelity enhancers such as the motion platform (e.g., Gray, 1980). Clearly, there is a large gap in our knowledge about the relationships among simulator physical fidelity, psychological fidelity, and training effectiveness. As long as these relationships are unknown, the tendency will be to design higher fidelity devices than are required or to build "state of the art" simulators.

The study discussed in this article provides a data point for the resolution of the fidelity requirement vs. tasks to be trained.

THE TRAINING DEVICE

In early 1978, two KC-135 navigators developed the concept and software to perform celestial training using a micro-computer. The concept was approved for evaluation by Headquarters, Strategic Air Command (SAC), and the actual feasibility of the concept was determined by the 4235th Strategic Training Squadron (STS) in August 1978. Final approval for the acquisition of the Celestial Training Devices (CTDs) was given in February 1979.

The 4235th STS was designated as the command program manager and given the responsibility for program implementation. CTD delivery started in August 1979 and was completed to all SAC operational squadrons by January 1980.

As presently configured, the CTD is simply a Northstar Horizon micro-computer with a 48K primary memory, a 5 1/4" floppy disk drive, an internal clock, a printer, and the associated software. The program is written in BASIC. Essentially any microcomputer with the required peripherals could be programmed as a CTD with only minor software modifications. The software was designed to enable a navigator to call up the computer program with very simple commands. The computer plays the part of pilot, aircraft, and celestial observer.

THE TRAINING PROFILE

The training profile is similar to actual celestial navigation in the aircraft with some exceptions. The navigator is in a classroom environment that lacks the realism of excessive noise, vibrations and other inconveniences of actual flying. Charts, logs, reference books and required equipment are available to be used simultaneously with the computer. The navigator follows the mission in real time, completing all routine work required on the charts and logs. Instead of coordinating with real people in the pilot and observer stations, he must communicate with the computer via the computer keyboard. Also, the computer has "built-in" errors that will cause the navigation to be in error, thus requiring continuous corrections.

PERFORMANCE MEASUREMENT

At the end of the simulated celestial navigation mission, the student or his instructor can take the Mission Summary data and replot the flight. This replot allows the student to compare his solution of the navigation mission with what the simulated aircraft, the computer, actually flew.

TRAINING EFFECTIVENESS STUDY DESIGN

When the celestial training device study was conceived in January 1980, most CTDs had already been delivered to the Air Force units. Since the Air Force routinely collects and files objective navigator evaluations in the form of in-flight performance skills on AF Form 157, it was then determined that a realistic study could be conducted. This study would compare in-flight navigation skill measurements, as recorded on the AF Form 157s before CTD practice, with similar skill measurements after the navigators had a chance to practice on the CTD.

The plan called for the flight record office at each base to send copies of all AF Form 157s for navigation flights prior to the beginning of CTD training to the 4235 STS. These forms were reviewed for completeness, and if they met predetermined criteria, they were used in this study.

The control group data consisted of the responses from the February 1980 request for all AF Form 157s prior to CTD training. This request yielded over 1,000 Form 157s. Initial review eliminated 700 of these that represented integrated flights because these flights allowed the navigator to use additional navigation aids such as inertial computers.

The 300 AF Form 157s that remained were reviewed for completeness. Certain omissions of less critical data points were considered acceptable. The data points considered to be less critical were: (1) final heading, (2) control time, and (3) Error (See below for definition of these terms). When these omissions were allowed, 121 of the Form 157s qualified for inclusion in the study and were used as control group data.

On May 28, 1980, CTDs had been in place long enough to justify acquiring the remaining data for the Celestial Training Device Study. A request went to all SAC squadrons using a CTD to provide the number of basic celestial navigation flights flown from Oct 1979 to May 1980. In addition, they were to send all AF Form 157s for basic navigation legs flown during the months of April and May 1980 to the 4235 STS.

Since the operational units understood that only basic celestial navigation flights were to be included, only 300 forms were received. These forms were reviewed under the same conditions and by the same person as the control group data. Ninety-nine of the AF Form 157s qualified and were used as the experimental data for our Celestial Training Device Study.

A total of twelve performance measures can be obtained from each AF Form 157. These are: (1) Final Heading, (2) Celestial Control Time, (3) Percent Reliability, (4) Computation Error, (5) Plotting Error, (6) Dead Reckoning (DR) Error, (7) Computation and Plotting Error, (8) Cumulative Error, (9) Rating, (10) Major Error, (11) Significant Error, (12) Minor Error.

ANALYSIS AND RESULTS

Of the twelve variables described earlier, all except instructor ratings yielded interval data (i.e., data for which an analysis of group means and variances would be appropriate). Instructor ratings were ordinal in nature; therefore, means analyses would not have been appropriate. Consequently, chi-square goodness-of-fit tests were performed on these data.

It was proposed at the outset of the study that data from B-52s and KC-135s should be combined into one group. This was felt to be valid since there is no logical reason to differentiate between navigators of the two aircraft with respect to the celestial navigation task. Procedures and training methods are essentially identical for both aircraft. Also, some of the Form 157s, which were to be included in the study, did not indicate which aircraft was involved. Therefore, a significant proportion of the data would have been excluded if it were necessary to discriminate between aircraft types within the data. To ensure the validity of this commonality assumption, the data for which the aircraft type was known were analyzed for potential differences. This ensured that no performance differences existed which were attributable strictly to aircraft type. Tables 1 and 2 present these comparisons between the pre-CTD and post-CTD groups, respectively, on the instructor ratings.

Chi-square goodness-of-fit tests conducted on these data yielded values of $\chi^2 = .771$ ($p > .5$) and $\chi^2 = .636$ ($p > .5$) for the data contained in tables 1 and 2, respectively. Table 3 presents this information for the cumulative error score. Two planned independent one-way ANOVAs were conducted to compare B-52 and KC-135 performance on the pre-CTD data $F(1,52) = .0076$ (NS); and on the post-CTD Data $F(1,52) = .34$ (NS). The analysis of these two critical measures of performance support the contention that B-52 and KC-135 data could be combined for analysis on an empirical, as well as a logical basis.

Table 4 presents means of the data comparing pre- and post- CTD performance with respect to the first nine variables. From this table, we can see that significant improvements were observed for Percent Reliability and Cumulative Error as a consequence of CTD training. No measures indicated a deterioration of performance. Table 5 presents the data for a chi-square goodness-of-fit test for the instructor rating data. The test statistic for this table is $\chi^2 = 9.4$ ($p < .025$). Table 6 presents the data for the major, significant, and minor errors to determine whether a shift had occurred in the type of errors made (i.e., relatively more minor and fewer major errors).

The χ^2 statistic for this table is $\chi^2 = 3.01$ ($p < .25$). However, also of potential interest was whether the mean number of each type of errors made per student had changed with the introduction of the CTD. Table 7 presents these means and indicates the test statistics and associated reliability of differences existing between the means of pre- vs. post-CTD training groups. It appears from Tables 6 and 7 that a reduction in the number of major errors may have occurred, although the effect was only significant at the .25 level.

Another question which had to be answered was how the number of in-flight celestial navigation legs had changed subsequent to the introduction of the CTD. The desirability of the CTD is very much a function of how large these reductions can be. SACR 51-52 and SACR 51-135 were modified so that the required number of celestial navigation legs were reduced. However, to ensure that this reduction had occurred, data were collected pertaining to the number of aircraft celestial navigation legs flown per month at each base. Some months had more flying days available due to the number of weekends and holidays. Consequently, these dates were adjusted to reflect the number of celestial navigation legs flown per available flying day.

Table 8 presents these data. To summarize, there was an average of 10.17 celestial navigation legs flown per available flying day at all of the bases studied during the months of October and November 1979, when the pre-CTD data were collected. There were an average of 7.43 celestial navigation legs flown per available flying day during the months of April and May 1980, when most of the post-CTD data were collected, a reduction of 3.27 celestial navigation legs per flying day. This represents a reduction of 30.6 percent. The number of navigators at these bases remained constant throughout the period of the study.

DISCUSSION

There is no doubt that the data obtained during this study support the contention that the CTD provides a highly effective training medium for the maintenance of celestial navigation skills. None of the measures showed a significant degradation of performance and several key measures (percent reliability, cumulative error, and instructor ratings) showed significant improvement at the 95% confidence level. Several other measures (number of major errors per flight, computation and plotting errors, and the distribution of error types) showed significant improvements at the 75% confidence level. The primary question was whether the substitution of the CTD for some inflight training had caused a deterioration of navigator performance. The answer is an unequivocal "No." In fact, we can surmise, with some degree of confidence, that in-flight performance has in some ways improved as a consequence of the CTD.

Because a performance improvement was observed, any estimate of the Transfer Effectiveness Ratio (TER) will probably underestimate the value of the CTD. However, using the most conservative data available, SACR 51-52, Volume 7, requiring one to three CTD flights per quarter, and the Transfer Effectiveness Ratio as defined by Roscoe (1971), we can estimate the TER = .33.

It is also interesting to compare the TER to the operating cost ratio of the B-52 and the KC-135 to the CTD. This provides us with an estimate of the cost effectiveness of the CTD. If we consider the operating costs to include only power (i.e., fuel and electricity) we find the following estimates:

Power Cost Estimates (per hour)

KC-135 (fuel) = 1560 gallons JP-4 @ \$1.17/gal = 1825.00
B-52 (fuel) = 2575 gallons JP-4 @ \$1.17/gal = 3013.00
CTD (electricity) = .5 KWH electricity @ \$.08/KWH = \$.04

or

KC-135/CTD Cost ratio = \$1825./\$.04
B-52/CTD Cost ratio = \$3013./\$.04

If the TER of .33 is an effective approximation of reality, the potential cost savings of the CTD are enormous.

COMMAND ACTIONS

HQ SAC, well aware of the training effectiveness and cost savings of the CTD, has already taken the following actions.

1. The CTD usage has been expanded to include initial qualification and instructor training as well as continuation training.
2. Regulations have been modified to permit the use of additional navigational aids allowed on inflight celestial navigation legs to include aircraft computers and inertial systems. This modification has reduced the number of pure inflight celestial navigation legs shown.

CONCLUSION

The Celestial Training Device represents a low-fidelity, in-house developed Part Task Trainer. It demonstrated an innovative use of existing computer technology. Its effectiveness lies not in the equipment related fidelity, but rather in procedural fidelity, i.e., the navigator had to perform actual required computing and plotting tasks. Realism was enhanced by the employment of "real time" and computer induced errors. This tends to indicate that scientists, educators, manufacturers and users of training simulators should be as concerned with the certification of simulator trained tasks, simulator training certification (SIMTRACERT), as they are with certifying the simulator's (physical) fidelity, commonly called Simulator Certification or SIMCERT.

Table 1				
Instructor Ratings - B-52 vs. KC-135, Pre CTD delivery				
	Highly Qualified (H)	Qualified (Q)	Qualified with Training (T)	Unqualified (U)
B-52	8 (57.1)	4 (28.6)	1 (7.1)	1 (7.1)
KC-135	28 (70.0)	8 (20.0)	2 (5.0)	2 (5.0)
$\chi^2 = .771$				
P = NS				

Table 2				
Instructor Ratings - B-52 vs. KC-135, Post CTD delivery				
	Highly Qualified (H)	Qualified (Q)	Qualified with Training (T)	Unqualified (U)
B-52	22 (71.0)	7 (22.6)	2 (6.5)	0 (0)
KC-135	67 (75.0)	18 (20.5)	3 (3.4)	0 (0)
$\chi^2 = .636$				
P = NS				

Table 3		
Cumulative Error Score - B-52 vs. KC-135		
	B-52	KC-135
Pre CTD	11.5	11.825
Post CTD	8.19	8.97
$F(1,52) = .0076$ P = NS		
$F(1,52) = .34$ P = NS		

Table 4					
Pre vs. Post CTD Celestial Navigation In-flight Performance					
Measure	Mean Pre CTD	Mean Post CTD	df	F _{1,df}	Significance Level
Final Heading	6.78	6.05	202	.24	NS
Celestial Control Time	1.66	1.70	104	.01	NS
Percent Reliability	97.57	99.76	207	7.55	<.01
Computation Error	.24	.32	209	.82	NS
Plotting Error	.53	.62	209	.62	NS
Dead Reckoning Error	2.07	2.04	207	1.1	NS
Computation and Plotting Error	6.21	4.93	207	1.42	<.25
Cumulative Error	11.24	8.79	207	4.53	<.04

Table 5					
Instructor Rating					
	Highly Qualified (H)	Qualified (Q)	Qualified with Training (T)	Unqualified	TOTAL
Pre CTD	63 (68.5)	18 (19.6)	4 (4.3)	7 (7.6)	92
Post CTD	89 (74.7)	20 (21.0)	5 (4.2)	0 (0)	119
NOTE: Percentages indicated in parentheses.					
$\chi^2 = 9.4$					
P = < .025					

Table 6				
Error Types				
	Major	Sig	Min	Total
Pre CTD	35 (28)	33 (26)	58 (46)	126
Post CTD	33 (20)	45 (27)	91 (54)	169
NOTE: Percentages in parentheses.				
$\chi^2 = 3.01$				
P = <.25				

Table 7					
Analysis of Major, Significant and Minor Errors made per flight					
Error Type	Mean-per CTD	Standard Deviation	Mean Post-CTD	Standard Deviation	t Significance Level
Major	.400	.875	.252	.585	1.14 <.25
Significant	.367	.775	.395	.693	.17 NS
Minor	.684	.522	.722	1.15	.69 NS

Table 8			
Number of Available Flying Days and Celestial Navigation Legs flown per Day by Month			
MONTH	FLYING DAYS	CELESTIAL NAV. LEGS	CELESTIAL NAV. LEGS/DAY
October 1979	23	267	11.61
November	20	193	9.65
December	17	177	10.41
January 1980	21	167	7.95
February	18	187	10.39
March	21	143	6.81
April	21	172	8.19
May	24	170	7.08

Use of Optical and Thermal Sights in Daylight Target Detection

Brian L. Kottas and David W. Bessemer
US Army Research Institute, Ft Knox Field Unit
Fort Knox, Kentucky 40121

Abstract

The Fort Knox Field Unit of ARI investigated the use of optical and thermal sights for daylight target detection. Armor soldiers were asked to detect targets in optical sight displays, thermal sight displays, or displays in which optical and thermal scenes alternated. In dense vegetation, alternating between thermal and optical sights produced more target detections than using either sight alone. In open terrain, using the optical sight produced the best performance. Overall performance with the thermal sight began poorly, but improved to the same level as the optical or alternating sight conditions in later trials. This research demonstrates the need for target detection training with the thermal sight; a marked increase in performance can occur when systematic feedback is provided.

Target detection occupies a critical place on the modern battlefield. The increased lethality of modern weapons demands rapid, effective target detection and identification. Thermal sights provided a quantum leap in our potential to detect and identify targets, but this technological advance must be coupled with an investigation of human perceptual capabilities in using them.

Soldiers clearly must choose thermal sights under conditions of low visibility. However, existing literature provides little empirically based guidance on incorporating its capabilities with those of optical sights to optimize target detection and identification during daylight with good visibility. This research focused on incorporating optical and thermal sight capabilities to maximize the number of targets detected and minimize the number of false alarms.

Past laboratory research has demonstrated that multiple observations can increase target detectability. Multiple, independent observations over time can be expected to increase the detectability of a target approximately as the square root of the number of observations (Swets, Shipley, McKey, and Green, 1964). The same results hold for independent observers attempting to detect targets. When their individual decisions are optimally combined, performance improves approximately as the square root of the number of independent observers (see Green and Swets, 1966). The sensory input from thermal and optical sights is analogous to the above examples. Input from each sight reflects a slightly different aspect, or dimension of the environment. One would hypothesize, on the basis of the results described above, that alternating views from optical and thermal sights should improve performance over using either sight separately.

Method. Armor soldiers in OSUT individually viewed slides containing from zero to three targets. Slides were taken either through an optical sight or a thermal sight. Observers pointed out targets in optical sight displays, thermal sight displays, or displays in which optical and thermal sight slides of the same scene alternated (the Alternating condition). Display times for all conditions were equated with a .4 second delay introduced between thermal and optical slides in the Alternating condition. Observers' response times were recorded for both correct detections and false alarms (mistakenly saying there was a target at a given location when, in fact, there was none). Search time on each trial was limited to 30 seconds. After each trial the experimenter pointed out correctly detected targets, missed targets, and areas that had been falsely identified as targets.

The Views, Opinions, and/or Findings Contained in this Report are those of the Author(s) and Should Not Be Construed as an Official Department of the Army Position, Policy, or Decision, Unless so Designated by Other Official Documentation.



Results and Discussion. The analysis focused on observers' first responses on each trial, since the first response is the most critical on the battlefield. For analysis, trials were divided into those on which slides contained only terrain with dense vegetation and those on which slides contained targets in open terrain (targets at the edges of dense vegetation or on bare and grassy areas).

Figure 1 shows performance for the three conditions in dense vegetation. Performance is described by the number of correct target detections divided by the total number of "target" responses. This measure takes into account both target hits and false alarms without the assumption of equal signal and noise distribution variances required by signal detection theory. Performance was graphed across pairs of trials in Figure 1 to smooth the curves. Comparison of the three sight conditions with a multiple t-test showed that the Alternating condition produced performance superior to that of both the thermal sight alone ($p < .05$) and the optical sight alone ($p < .001$). The difference between the optical and thermal sights alone was insignificant.

Figure 1 further suggests that with practice observers will become more proficient in thermal target detection. Further research is required to determine the time course of skill acquisition and performance asymptotes, to determine training requirements with the thermal sight.

Analysis of response times showed that the increased accuracy in the Alternating sight condition occurred at the expense of time; responses in the Alternating sight condition were significantly slower than in either of the other two sight conditions. Table 1 provides an indication of the speed/accuracy tradeoff in dense vegetation.

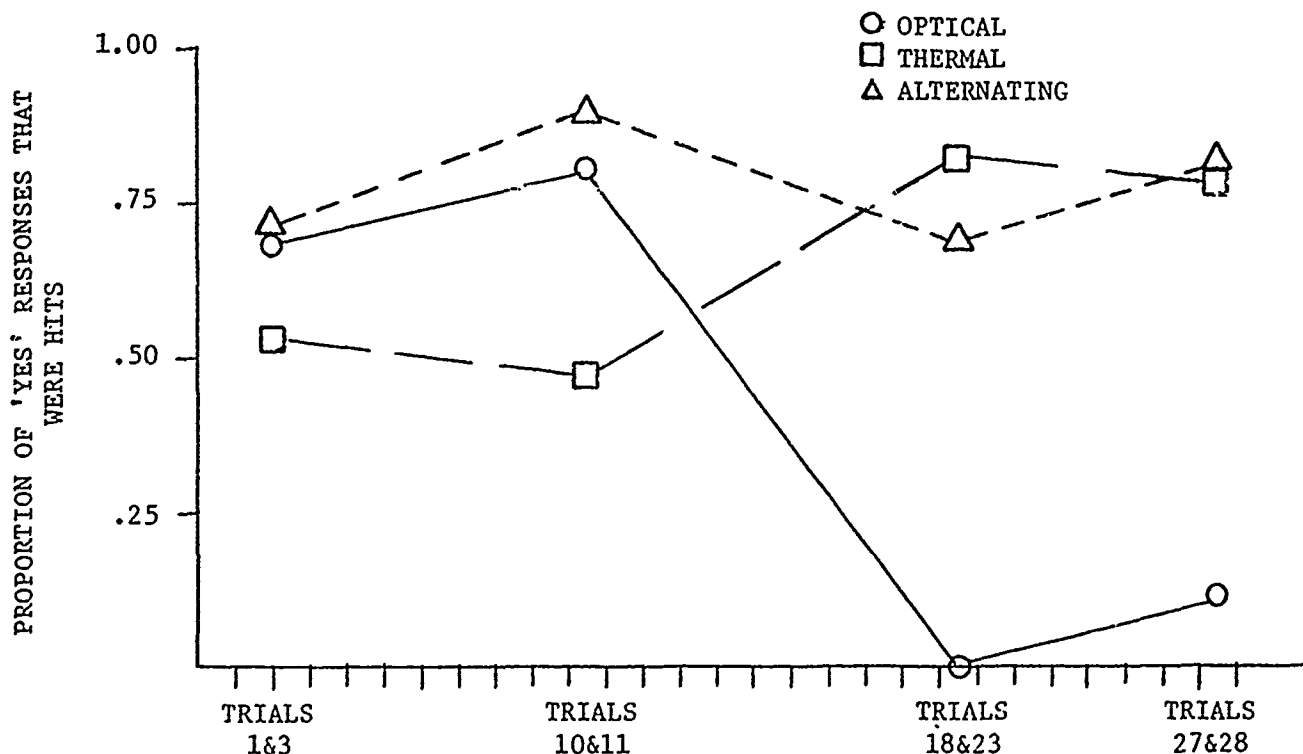


Figure 1. Proportion of 'Yes' Responses That Were Correct Across Pairs of Trials in Dense Vegetation.

TABLE 1

PERFORMANCE (DENSE VEGETATION)	SIGHT CONDITION		
	Optical	Thermal	Alternating
Proportion of Trials			
First Response Hits	.22	.47	.55
First Response False Alarms	.27	.24	.14
Average Detection Time (sec)	9.66	12.36	15.31

The practical implications of the speed/accuracy tradeoff depend on the tactical situation; in some situations speed is the only consideration. However, from a covered and concealed defensive position, it would seem that speed in detecting targets is of little value if over three-fourths of the targets are missed, as was the case with the optical sight condition. For many engagements in dense vegetation, it may be more important to detect as many targets as possible to optimally allocate fires before revealing your position.

Performance in detecting targets in open terrain yields a slightly different picture. In open terrain, there was no significant difference between optical and Alternating sight conditions in terms of proportion of targets detected. Detection performance with the thermal sight alone was significantly poorer than in the other two conditions, but showed signs of improving over practice. Figure 2 shows these results graphically.

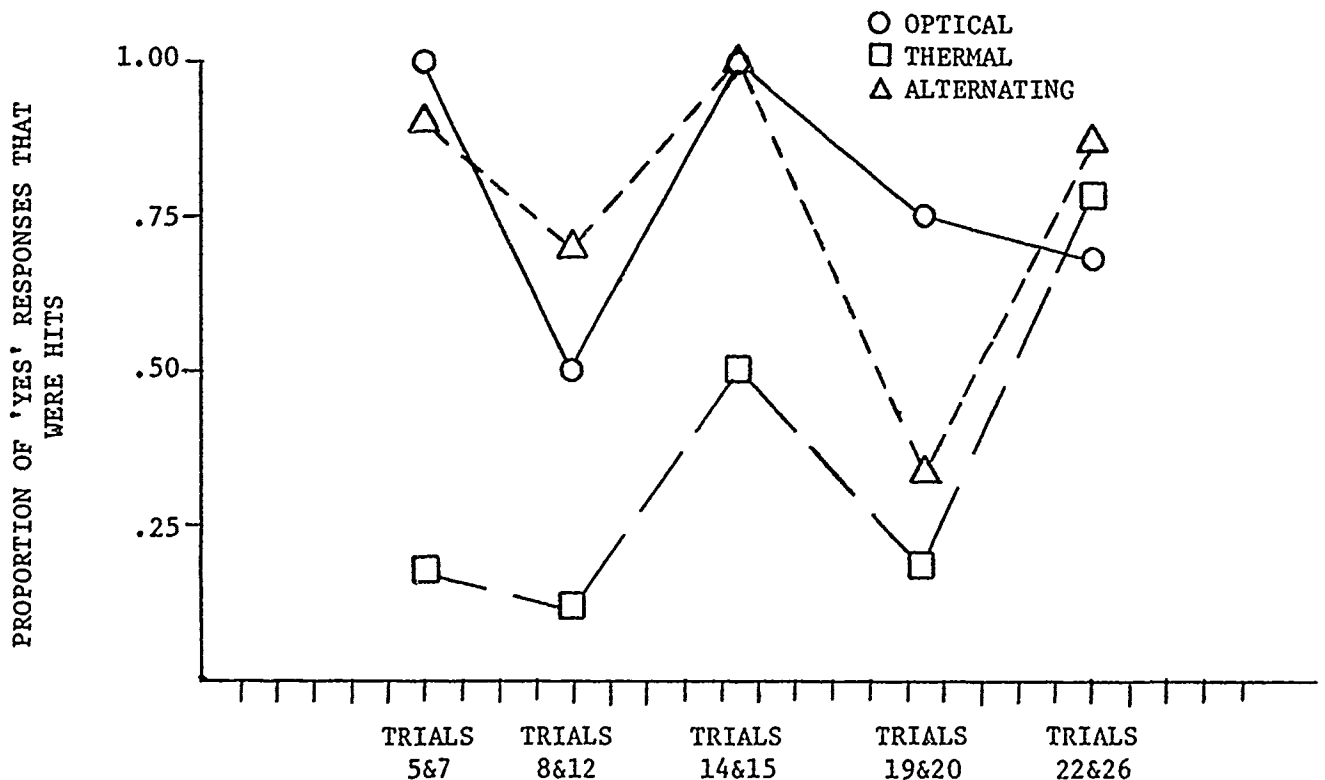


Figure 2. Proportion of 'Yes' Responses That Were Correct Across Pairs of Trials in Mixed Terrain.

Analysis of response times in open terrain revealed that correct detections with the optical sight were significantly faster than with the other two sighting conditions. Response times for false alarms showed no significant differences. Table 2 shows the speed/accuracy tradeoff for open terrain. Considering the improvement in performance with the thermal sight alone over trials, it would be interesting to evaluate the average proportion of targets detected per second of search with the thermal sight after more training.

TABLE 2

<u>PERFORMANCE (OPEN TERRAIN)</u>	<u>SIGHT CONDITION</u>		
	Optical	Thermal	Alternating
Proportion of Trials			
First Response Hits	.58	.24	.61
First Response False Alarms	.14	.41	.14
Average Detection Time (sec)	6.78	11.29	11.47

The data are consistent with the information cited above from Swets et al. that target detectability increases as the square root of the number of independent observations. Data from the Alternating sight condition exceeded the expected average of the other two sight conditions by approximately $\sqrt{2}$. The data also revealed that the probability of false alarms in the Alternating sight condition was lower than the average expected false alarm probability for the other two sight conditions by approximately $1/\sqrt{2}$. Table 3 shows some values along the regression lines that illustrate these findings.

TABLE 3

<u>DENSE VEGETATION</u>		<u>TRIAL BLOCK</u>			
Proportion of Trials	1st Qtr	2nd Qtr	3rd Qtr	4th Qtr	Mean
First Response Hits					
Alternating Optical/Thermal	.58	.54	.51	.47	.52
Predicted (Average X $\sqrt{2}$)	.51	.50	.49	.48	.50
First Response False Alarms					
Alternating Optical/Thermal	.18	.16	.15	.13	.16
Predicted (Average X $\sqrt{2}$)	.16	.17	.18	.20	.18

<u>OPEN TERRAIN</u>		<u>TRIAL BLOCK</u>			
Proportion of Trials	1st Qtr	2nd Qtr	3rd Qtr	4th Qtr	Mean
First Response Hits					
Alternating Optical/Thermal	.87	.71	.55	.38	.63
Predicted (Average X $\sqrt{2}$)	.69	.62	.55	.48	.58
First Response False Alarms					
Alternating Optical/Thermal	.14	.14	.14	.14	.14
Predicted (Average X $\sqrt{2}$)	.25	.22	.18	.15	.20

Although space does not permit a full discussion here, analysis of the data showed an extremely high proportion of first response false alarms with the thermal sight on early trials in open terrain. Later trials showed a dramatic decrease in false alarms. It is unclear whether the decrease was due to increased discrimination of targets from non-targets, or whether observers learned the location of hot spots on the terrain that looked much like targets, since the slides were taken in a fairly restricted area of terrain. Probably, some combination of learning about the terrain and learning to discriminate targets from non-targets was responsible for the increase in thermal target detection performance. However, further research must be done to assess the relative contributions of these two factors.

A further conclusion that can be drawn from these data is that detecting targets using thermal sights is not automatic; performance improved over practice. Simply allowing someone to look through a thermal sight for a while will almost certainly fail to produce acceptable levels of target detection performance. It should be relatively easy to develop simulated thermal target detection training displays; since the output of the thermal sight is a video display, the likelihood of developing transferable training for use in a classroom setting using currently available technology seems excellent. This research points out one caveat in developing classroom thermal target detection training -- since there may be some terrain-specific learning, the target scenes used should have a number of different backgrounds to provide varied experience in discriminating targets from terrain features that have a high probability of falsely being identified as targets.

The immediate relevance of this research for combat, given the likelihood of terrain-specific learning, is that when setting up in a defensive position, TCs and gunners should examine the terrain through both the thermal and optical sights. This will allow them to locate hot spots that are potential opportunities for false alarms to occur.

REFERENCES

- Green, D. M. and Swets, J. A. Signal Detection Theory and Psychophysics. New York: Wiley, 1966.
- Swets, J. A., Shipley, E. F., McKey, M. J., and Green, D. M. Multiple observations of signals in noise. In J. A. Swets (Ed.) Signal Detection and Recognition by Human Observers. New York: Wiley, 1964.

ABSTRACT

This study tested a training concept termed reverse-role training. Reverse-role training involves practice by an individual on tasks which he does not perform in the operational environment but his operational performance is dependent upon others performing these tasks. This can be best perceived in terms of game playing, such as air-to-air combat, whereby the player will perceive practice playing the opponent's position (e.g., American fighter pilots receiving training in Soviet aircraft). A study was conducted in which 32 subjects were trained to play a video air-to-air combat game with each "aircraft" having different performance characteristics. Results indicated significant differences on a number of strategy measures and a probable interaction with aircraft type on the number of wins, losses, and draws.

by K. Ronald Laughery, Jr. and
Jan L. Ditzian

Calspan Advanced Technology Center
P.O. Box 400
Buffalo, New York 14225

Introduction

The advent of modern simulation technology has created opportunities for many new and exciting approaches to training which have not yet been realized. Until recently, the use of simulation in training has been solely for the purpose of "simulating" operation of the actual equipment in a realistic environment. Even with the advanced instructional features which are available on many existing simulators, most are going underutilized due to a lack of understanding of how they can best be applied (Isley and Miller, 1976). Given that simulators have capabilities with respect to training which in many ways exceed that of the operational equipment they are simulating, one of the priorities of the training research community should be to determine the means by which simulation training can be used to improve training programs with respect to cost-effectiveness and/or improvement of the outcoming trainees' capabilities. Research programs are underway within the Department of Defense to explore concepts such as backward chaining (Bailey, Hughes, and Jones, 1979) and automated adaptive training (Chaffield and Gidcumb, 1977, Feurzeig, Cohen, Lukas, and Schiff, 1975). The proposed research effort is intended to be another step towards the full realization of the capabilities of simulation training.

Specifically, the proposed experiment explored the application to simulation training of a training concept recently advanced by Klein (1977). In this report, Klein identifies and discusses a concept he terms the "phenomenological approach to training". Basically, the concept states that there are components of training which cannot be identified in an algorithmic fashion (e.g., task analysis). In order for adequate training, a student must experience the phenomena directly or indirectly. Klein states that "the whole is greater than the sum of its parts" and, therefore, learning will not be adequately realized without this "phenomenological training". Analogies of tennis, bicycle riding, and flying airplanes sufficiently illustrate the concept. The study discussed tested a training approach termed reverse-role training, in which simulators can be used to provide a phenomenological experience which would otherwise not be feasible.

Reverse-role training can be defined as that pertaining to the practice of complimentary tasks which are performed not by the trainee in the operational environment but by someone else. Obviously, for this training to be of utility, there must be some interdependence between the complimentary tasks practiced and the tasks performed by the trainee in the operational environment. This task interdependence can be either of an adversary nature (e.g., practicing the piloting of the other aircraft in an air-to-air combat situation) or involve mutual coordination (e.g., practicing boom-operation in the training of piloting the receiver aircraft during aerial refueling). Additionally, reversal-role training should be superior to cognitive training of the complimentary tasks in that the phenomenological training should be superior to non-experiential approaches. The question which the study attempted to answer was whether reversal-role training of the adversary's task results in improved operator performance in the actual gaming situation.

Method

The Task - A video air-to-air combat game was developed for an Apple II microcomputer. The game was a two-dimensional, one-on-one game. The subject did not control altitude, only speed and turn rate. The subject's perspective was as though he were outside and above the aircraft looking down upon the gaming area. An example of the display he monitored is presented in Figure 1. He could always see the location of both his and the opponent's aircraft.

The subject had three controls: a stick, throttle and firing button. The stick controlled rate of direction change. The throttle controlled the speed of the aircraft, similar to an accelerator on an automobile. The throttle also controlled the fuel consumption rate. The firing button controlled the player's weapons. When a subject wanted to fire at the opponent, he depressed the firing button. Each player was limited with respect to fuel availability and the number of weapons which could be fired during any game. Each game was completed by either player achieving a hit on the opponent or both players exhausting either fuel or weapons.

The aircraft differed significantly with respect to handling and weapons characteristics. A summary of these differences is presented below:

<u>Characteristic</u>	<u>Aircraft 1</u>	<u>Aircraft 2</u>
Maximum speed	Lower	Higher
Maximum turn rate	Higher	Lower
Acceleration	Faster	Slower
Weapons range	Longer	Shorter
Weapons accuracy requirements	More accuracy required	Less accuracy required
Number of weapons	3	5
Available fuel	12,000 units	15,000 units

A pilot study was conducted to ensure that the aircraft were approximately equal overall, yet they required the use of different strategies. After some manipulation of each aircraft's characteristics, this was reasonably confirmed.

Subjects - Thirty-two male subjects between the ages of 17 and 28 were used in this study. They were recruited by signs placed in several locations at the State University of New York at Buffalo. Subjects were paid \$25.00 for their participation in the experiment. All subjects who began in the experiment completed the experiment (i.e., no subject attrition).

Experimental Procedure - Subjects were assigned to one of four conditions based on a 2X2 factorial design. The first factor was aircraft. Subjects were assigned to either Aircraft 1 (AC1) or Aircraft 2 (AC2). The second factor was training type. Subjects received either Reverse-Role Training (RRT) or Standard Training (ST). Eight subjects were randomly assigned to each of the four groups (i.e., AC1-RRT, AC1-ST, AC2-RRT, and AC2-ST).

All subjects received a total of 300 training trials (games) over a two day period. On the first day, subjects received a verbal briefing on the

instructions and desirable tactics for both aircraft. Then, all subjects received 120 training trials in their primary aircraft. On the second day, subjects in the RRT group received 60 trials in their primary aircraft and then switched to the other aircraft for 80 trials. This training on the "opponent's" aircraft was reverse-role training. Finally, they received another 40 trials on their primary aircraft. Subjects in the ST group received 180 trials in their primary aircraft only. The pilot study indicated that these training intervals were adequate for near asymptotic performance by subjects.

All training trials were played against an "instructor pilot". One instructor pilot was assigned to each aircraft so subjects receiving training in Aircraft 1 played against the instructor from Aircraft 2 and vice versa. Prior to commencing the experiment, the instructors played approximately 3,000 games against one another to minimize potential instructor learning effects over the course of the study.

After training, each subject played against two other subjects on a third day. Both opponents were from the other aircraft (e.g., AC1 subjects played two AC2 subjects) but one of the opponents was Reverse-Role Trained (RRT) and one was Standard Trained (ST). The subject played against each opponent for 100 trials.

Results

Strategy Analysis - Data were collected on each subject during the training and experimental sessions so that all trials could be completely replayed for a detailed strategy analysis. It was of interest to determine how player strategy was affected by Reverse-Role training. Based on protocol analysis of the instructors, 39 measures of strategy were developed. These measures were evaluated during the course of each trial by examining stick and throttle positions and by examining the geometry of the aircraft relative to one another. These included measures of aircraft speed, acceleration, turn rate, pointing direction relative to the opponent, head-on encounters, intercepts, attacks from the side, fuel usage rate, and missed opportunities. For a more detailed discussion of these measures and how they were derived, see Laughery (1981).

A 2X2X2 ANOVA was performed on each of these measures for the 200 experimental trials (i.e., against other subjects). The first two factors were between subjects, aircraft type and training type, and the third factor was within subjects, opponent training type. Each subject played against opponents from both training types (RRT and ST). Of these 39 measures, 9 showed significant differences with respect to training type at the $p < .10$ level or better. These are the following:

Measure Name	Description	Nature of Difference	p
Positive Acceleration	% of trial during which subject was accelerating.	RRT < ST	.10
Towards the Opponent	% of trial during which subject's aircraft was getting closer to the opponent.	RRT > ST	.10
Proximal Pointing Towards	% of trial during which subject was pointing towards the opponent's aircraft.	RRT > ST	.05
Proximal Pointing Orthogonal	% of trial during which subject was pointing neither towards nor away from opponent.	RRT < ST	.10
Proximal Pointing Away	% of trial during which subject was pointing in the opposite direction of opponent.	RRT < ST	.10
Relative Pointing Same	% of trial during which both aircraft were pointing in the same direction.	RRT < ST	.01
Relative Pointing Opposite	% of trial during which both aircraft were pointing in opposite directions.	RRT > ST	.01
Change in Proximal Pointing - Away	% of trial during which subjects is turning away from opponent.	RRT < ST	.10
Meet -	% of trial during which player was attempting to meet the opponent.	RRT > ST	.05

Eight variables were found to be significant at the $p < .10$ level or better with respect to opponent training type. These differences tended to support the findings with respect to subject training type. For the aircraft factor, 26 of the 39 variables showed significant effects, 24 of which were at the $p .05$ level or better. This supported the contention that the aircraft were fundamentally different, an essential requirement if reverse-role training is to be useful.

Wins/Losses/Draws Analysis - To determine whether reverse-role training had impact upon the "bottom-line" of performance, chi-square analyses were performed on the distribution of wins, losses, and draws as a function of

training type. Table 1 presents the total wins for RRT and ST subjects when they were in head-to-head competition (i.e., trials involving RRT vs. RRT or ST vs. ST subjects were not included).

Table 1
Wins for Training Conditions
in Head-to-Head Competition

<u>Condition</u>	<u># Wins</u>
RRT	717
ST	735

The Chi-Square test statistic for these data were $\chi^2 = .22$ ($p > .5$), indicating that no main effect of training had occurred.

The data regarding aircraft type are presented in Table 2. The Chi-Square test statistic for these data is $\chi^2 = 25.5$ ($p < .001$) indicating

Table 2
Wins for Aircraft Conditions

<u>Condition</u>	<u># Wins</u>
Aircraft 1	1,304
Aircraft 2	1,575

321 Draws

substantial aircraft differences. The question which then arose was whether there might be an interaction of training type with aircraft type such that subjects from one aircraft would benefit from RRT and the subjects from the other aircraft would be hurt by RRT. To test this, RRT subjects from each aircraft playing against ST opponents from the opponent were compared against ST vs. ST match-ups. The assumption was that, given common opponent training types, we could test the effect of subject training type within each aircraft condition. These data are presented in Tables 3 and 4 for Aircraft 1 and 2, respectively.

Table 3

Aircraft 1 - RRT vs. ST

	Wins	Losses	Draws
RRT-AC1 vs. ST-AC2	323	396	81
ST-AC2 vs. ST-AC2	288	416	96

Table 4

Aircraft 2 - RRT vs. ST

	Wins	Losses	Draws
ST-AC1 vs. RRT-AC2	394	339	67
ST-AC1 vs. ST-AC2	416	288	96

Chi-Square test statistics for tables 3 and 4 are $\chi^2 = 3.77$ ($p < .16$) and $\chi^2 = 9.91$ ($p < .01$), respectively. This indicates that subjects from Aircraft 1 may have been helped by reverse-role training while subjects from Aircraft 2 were hurt by it.

Discussion

Reverse-role training clearly seems to influence human performance. This was manifested in both strategy and win/loss analyses. The strategy analysis indicated changes in variables which seemed to point towards increased aggressiveness on the part of reverse-role trained subjects. They were more likely to be moving and pointing towards the opponent, and they were more likely to be attempting to meet the opponent. All of these measures indicate an increased likelihood of reverse-role trained players to be seeking a confrontation or, otherwise stated, expressing aggressive behavior. However, being aggressive is not necessarily beneficial. One should only be aggressive if one has the skills necessary to win. The utility of reverse-role training based upon the increase in aggressive playing behavior should be judged on a case-by-case basis. Moreover, further research should be conducted to support this finding and determine its generality to other tasks.

The ultimate utility of reverse-role training was not clearly supported by the analysis of wins, losses, and draws. The apparent interaction with aircraft type leads to more questions than answers. It is hypothesized that the ultimate utility of reverse-role training will be mediated by 1) the total amount of training required for the primary task, 2) when reverse-role training is delivered during training program, and 3) how much reverse-role training is received. In this experiment we looked at two tasks (i.e., flying two

different aircraft) under the assumptions that total training requirements were the same and the application of reverse-role training should be applied in a similar manner. These assumptions may not be valid. In that event, we would expect different effects of reverse-role training. Based upon this study, we would have to say that reverse-role training should be applied cautiously only after exploring different amounts of reverse-role training given at different points in the training program. Future parametric studies where these variables are systematically varied and their effects studied may clear up this issue.

Clearly, reverse-role training offers potential. If we can provide reverse-role training through simulation, we may have a cost-effective way of facilitating improved human performance through experience in a unique training environment.

References

1. Isley, R.N. & Miller, E.J., "The Role of Automated Training in Future Army Flight Simulators". HUMRRO Report No. HUMMRO-FR-ED-76-27, Human Resources Research Organization, Alexandria, Virginia, October 1976, AD-A034 414.
2. Chatfield, D.C. & Gidcumb, C.F., "Optimization Techniques for Automated Adaptive Training Systems". Report No. NAUTRAEQUIPC-77-M-0575, Naval Training Equipment Center, Orlando, Florida. July 1977, AD-A052-631.
3. Feurzeig, U., Cohen, D., Lukas, G., & Schiff, M., "Research on High-Level Adaptive Training Systems". Report No. NAUTRAEQUIPC-74-C-0081-6, Naval Training Equipment Center, Orlando, Florida. November 1975, AD-A022-291.
4. Klein, G.A., "Phenomenological Approach to Training". Report No. AFHRL-TR-77-42, Air Force Human Resources Laboratory, Brooks AFB, Texas. August 1977, AD-A043-920.
5. Laughery, K.R., "An Evaluation of Reverse-Role Training in Strategic Game Playing". Unpublished Doctoral Dissertation, 1981.

The Effects of Human Relations Training on the Leadership Qualities of Military Personnel

William M. Barkley
Peabody College of Vanderbilt University

Richard L. Percy
Peabody College of Vanderbilt University

Abstract

This study attempted to determine the effects of a university graduate degree program in Human Development Counseling on the management and leadership qualities of selected U.S. Air Force personnel stationed in Europe. More specifically, the effects of such training on two dimensions of Leadership--Consideration of Others and Concern for Structure were investigated. The results confirmed the hypotheses that such training significantly enhances the participants' consideration for others without lowering their concern for structure. Such results argue for the use of human relations training models as appropriate for the development of leadership traits among military personnel.

Middle and upper level Air Force managers are continually being evaluated on their ability to not only accomplish the general and specific goals of a military organization but also on their ability to establish relationships with their subordinates which are characterized by mutual trust, respect for ideas and consideration of feelings (Cherry, 1975). This emphasis on concern for the individual as an integral part of goal attainment is highlighted by the recently published study on families in the Air Force (Orthner, 1980). It is becoming increasingly clear that attention must be given to "people" in order to deal effectively with retention of experienced personnel, moral, disruptive behavior and mission related performance.

In August, 1975, George Peabody College for Teachers was invited by the U.S. Air Force in Europe to establish overseas graduate degree programs in Human Development Counseling (OHDC) for military personnel and their dependents in the United Kingdom and in West Berlin. Approximately 50% of the military members enrolled in the OHDC program had no immediate intention of becoming professional counselors. Instead, they intended to use the knowledge and skills gained through the program to enhance their supervisory skills in the military as well as to enhance their promotability.

Numerous authors (Hall, 1970; Blake & Mounton, 1964; Gordon, 1977) have demonstrated the relevance of human relations training for individuals with management and supervisory responsibilities.

Cherry (1975) discovered that the cluster of traits having to do with warmth, capacity for close interpersonal relationships, compassion and considerateness correlated significantly with the qualities of productivity, creativity, cooperativeness, and job satisfaction. More recently, Leonard (1981) discusses the growing trend among corporations to provide social and psychological services to their employees as a means of improving profits through better care of human resources. Because the competencies for the Peabody overseas program were consistent with those trends in management, it seems logical to believe that those military members who have participated in OHDC have actually been developing attitudes, values and interpersonal skills necessary to improve their effectiveness as managers and supervisors with the military. The critical nature of a supervisor's job in the military makes the development of consideration for individuals at the expense of getting the job done unacceptable. Thus, the relative effect of the OHDC program on the military population needs to take into account the extent to which it affected the tendency to structure supervisory roles toward goal attainment.

The use of good human relations skills on the job with no loss in concern for accomplishing the mission is predicated on appropriate attitudes and values. As Mahon and Altman (1977) have pointed out, "... to what extent is it really possible to develop skills not entirely consistent with such attitudes and values" (p. 43). Thus, the Leadership Opinion Questionnaire (LOQ) developed and revised by Fleishman (1951 & 1969) was selected as the criterion measure for seeking answers to the question: Do military members, as a result of educational training through a university based program in Human Development Counseling, significantly increase their consideration of people without adversely affecting their concern for goal attainment?

Instrumentation

The Leadership Opinion Questionnaire (LOQ) was designed by Fleishman (1960) to provide measures of two important dimensions of supervisory leadership - Consideration of Others (C) and Concern for Structure (S). The LOQ has been used with a variety of organizations, including the Air Force, as a means of discriminating between high and low achieving managers (Halpin, 1954; Oaklander & Fleishman, 1964). The split-half reliability coefficients for various samples ranged from .62 to .89 for C and from .69 to .88 for S. Test-retest reliability data were collected from first line executive and Air Force NCO's. Coefficients ranged from .77 to .80 for C and from .67 to .74 for S (Fleishman, 1969).

Empirical validity studies carried out in various organizations with different type managers and supervisors provided significant correlations between the LOQ scales and a variety of different

criteria. While the pattern was not universal, no cases were found where low C goes with good job performance. Thus, low C scores are more often indicative of an undesirable supervisory situation. While the results for S depend more on the nature of the setting, the above average C & S pattern appears to be most clearly associated with effective managerial style (Fleishman, 1969).

Procedures

In February, 1980, the LOQ was administered to all 74 OHDC students attending class at two arbitrarily selected U.S. Air Force bases in the United Kingdom. In April, 1980, the LOQ was administered to all 12 military members attending an OHDC class at a third base. In addition, each of the 12 students was asked to identify an individual who was not enrolled in the OHDC program who was of equal rank, held a similar job and who was of the same sex as the student. The LOQ was then administered to these non-OHDC military members by the persons identifying them.

In the February administration there were four subjects who turned in incomplete instruments and in April one subject did not turn in a complete instrument. These subjects were eliminated from the study.

During both administrations, subjects were asked to provide demographic data and to indicate the number of credit hours completed in the 30 semester hour OHDC program as of the beginning of the current quarter. The 13 OHDC subjects in the February administration who indicated that they had completed more than nine but less than 21 credit hours were eliminated from the study. A 2 x 2 least squares analysis of variance was computed for subjects who had completed 0-9 versus 21+ credit hours by military versus civilian status. In addition, a priori comparisons were made between military members having completed 0-9 hours and those having completed 21 or more hours in the program. A t-test for correlated samples was computed on the 11 pairs of subjects from the April administration.

One hundred percent of the military members participating in the study were officers or senior non-commissioned officers. All 19 of the civilians were female in comparison to only eight out of the 60 OHDC military members.

Results

This study was designed to test two general hypotheses:

1. Participation in the OHDC program would be associated with higher scores on the Consideration of the Individual (C) scale of the LOQ.

2. Participation in the OHDC program would not be associated with lower scores on the Concern for Structure (S) scale of the LOQ.

Table 1 presents the means and standard deviations for scores on the two dependent variables broken out by military versus civilian status and number of semester hours completed for the February administration. Tables 2 and 3 present the least squares analysis of variance summary tables for C and S, respectively.

Tables 2 and 3 indicate that the only significant result is a main effect due to number of hours in the program with respect to Consideration of the Individual ($F = 8.47 > F_{.95} (1,53) = 4.03$). From inspection of Table 1 it can be seen that cell means are in the predicted direction. These results are taken as support for both hypotheses.

Insert Tables 1, 2 and 3 about here

Since the primary focus of this study is on students who are members of the military, a priori comparisons were made on both dependent variables between military students with 0-9 hours versus those with 21+ hours in the program. The results again support both hypotheses. The comparison with respect to C yielded $F = 68.31 (> F_{.95} (1,53) = 4.03)$, while $F = 0.06 (< F_{.95} (1,53) = 4.03)$ was obtained with respect to S.

Further support for the hypotheses was generated from 12 military OHDC students matched with non-OHDC students. Table 4 presents the means and standard deviations for both dependent variables. The t-test for correlated samples yielded $t = 1.96 (> t_{.95} (10) = 1.81)$ for C and $t = 0.74 (> t_{.95} (10) = -1.81)$ for S.

Insert Table 4 about here

In all three tests the null hypotheses regarding Consideration of Others were rejected at the .05 level of significance. Conversely, the data consistently failed to reject the null hypotheses with respect to Concern for Structure. In addition, a post hoc decision was made to compute Pearson Product Moment Correlation Coefficients between C and S scores.

Table 5 shows that while significant negative correlations were found for civilians with over 21 hours, members of the military not enrolled in the OHDC program and for the total sample, significant correlations were not obtained for any group of military members who were enrolled in the OHDC program.

Insert Table 5 about here

Discussion

The results of this study offer support for the hypotheses that participation in the OHDC program is associated with the development of Consideration of the Individual. Moderate support is evident for the hypothesis that participation in the OHDC program is not associated with lower Concern for Structure. Caution with respect to the latter conclusion must be exercised due to the low statistical power of the various statistical tests employed. Power hovered around .10 with respect to small effect sizes (Cohen, 1977). On the other hand, the consistent confirmations of a priori predictions can be taken as meaningful support for both hypotheses.

The most serious factor jeopardizing internal validity is subject selection. It is certainly a plausible hypothesis that there could exist a selection by treatment interaction resulting in positive results only for military members who would choose to enter a program such as OHDC. In addition, the civilian/military status variable was confounded with sex of subject.

Conclusions

It seems clear that military members who chose to enter a Human Development Counseling program were more concerned with the individual than their counterparts who did not and that students with 21+ hours in the program were more concerned with individuals than those just starting the program. Furthermore, there is no evidence that consideration of the individual is related to less concern for accomplishing the mission of the Air Force for military members enrolled in the OHDC program. The authors would like to draw the conclusion that participation in OHDC type programs results in positive outcomes for Air Force personnel in general. However, such a conclusion goes beyond the data. It remains to be seen whether an OHDC type program would have any benefit for Air Force personnel in general.

References

- Blake, R. R. & Mouton, J.S. The managerial grid, Houston: Gulf Publishing Co., 1964.
- Cherry, G.W. The serendipity of the full functioning manager. Sloan School of Management, Massachusetts Institute of Technology, 1975 (Unpublished manuscript).
- Cohen, J. Statistical power analysis for the behavioral sciences, New York: Academic Press, Inc., 1977.
- Fleishman, E.A. Leadership opinion questionnaire, Chicago: Science Research Associates, Inc., 1960.
- Fleishman, E.A. Leadership opinion questionnaire, Chicago: Science Research Associates, 1969.

- Gordon, T. Leadership effectiveness training, New York: Bantam Book, 1977.
- Hall, J. What makes a manager good, bad, or average? Psychology Today, 1976, 3, 52-55.
- Halpin, A.W. The leadership ideology of educational administrators and air craft commanders. Harvard Educational Review, 1955, 25, 18-32.
- Leonards, J.T. Corporate psychology: An answer to occupational mental health. Personnel and Guidance Journal, 1981, 1, 47-51.
- Mahon, B. R. & Altman, H.A. Skill training: Cautions and recommendations. Counselor Education and Supervision, 1977, 1, 42-50.
- Oaklander, H. & Fleishman, E.A. Pattern of leadership related to organizational stress in hospital settings. Administrative Science Quarterly, 1964, 8, 520-532.
- Orthner, D.K. Families in blue: A study of married and single parent families in the U.S. Air Force, Greensboro, N.C.: Family Research & Analysis, Inc., 1980 (USAF Contract #F33600-79-C-0423).

Table 1

Descriptive Data for Concern for Individual and for
Structure by Status and Number of Hours in HDC

Status	No. Hrs.	Consideration of the Individual			Concern for Structure		
		0-9	21+	Total	0-9	21+	Total
Military	n	25	13	38	25	13	38
	\bar{x}	53.68	59.23	55.58	46.24	44.38	45.61
	s	5.76	6.07		5.63	9.71	
Civilian	n	10	9	19	10	9	19
	\bar{x}	57.60	60.44	58.95	46.10	43.56	44.89
	s	6.08	4.72		7.65	7.13	
Total	n	35	22	57	35	22	57
	\bar{x}	54.80	59.73	56.70	46.20	44.05	45.37

Table 2

Summary Table for Consideration of the Individual

Source	SS	df	M.S.	F
A(Status)	95.22	1	95.22	2.89
B(Hours)	279.47	1	279.47	8.47*
AB	22.37	1	22.37	.68
Error	1748.37	53	32.99	

*F_{.95} (1,53) = 4.03

Table 3

Summary Table for Concern for Structure

Source	SS	df	M.S.	F
A(Status)	2.38	1	2.38	0.04
B (Hours)	58.70	1	58.70	1.10
AB	1.41	1	1.41	0.03
Error	2824.76	53	53.30	

$$F_{.95} (,53) = 4.03$$

Table 4

Description Data for Consideration of the Individual and
Concern for Structure for Matched Pairs by HDC students
and Non-HDC Students

Consideration of the Individual Concern for Structure

	OHDC	Non-OHDC	OHDC	Non-OHDC
n	11	11	11	11
\bar{x}	56.09	53.55	43.91	46.00
s	4.06	4.08	6.80	11.19

Table 5

Correlations between Concern for the Individual
and Concern for Structure for all Groups used in
This Study

	Group	<u>n</u>	<u>r</u>	<u>t</u>
February Administration	Civilians 0-9	10	-.42	-1.31
	Civilians 21+	9	-.70	-2.59
	Military 0-9	25	-.04	- .19
	Military 21+	13	-.31	-1.14
April Administration	Military Peabody OHDC	11	-.25	- .77
	Military Non-Peabody Non-OHDC	11	-.53	-1.87**
Total Sample		79	-.32	-2.96***

* $t_{.95} (7) \leq - 1.89$

** $t_{.95} (9) \leq - 1.83$

*** $t_{.95} (77) \leq - 1.67$

USE OF WEAPONER MARKSMANSHIP TRAINER IN PREDICTING
M16A1 QUALIFICATION PERFORMANCE

Joel D. Schendel and Frederick H. Heller

U.S. Army Research Institute Field Unit
Fort Benning, Georgia

Most scenarios for a full-scale confrontation between the U.S. and any of its major potential adversaries suggest that the majority of Army units will have to be prepared to fight immediately with little or no opportunity for post-mobilization training. Comparisons of the military strengths of the U.S. and the Warsaw pact countries also indicate that the U.S. forces are likely to be heavily outnumbered, often by a ratio of five to one or more. To have any chance for success, the Army will have to maintain a consistently high level of combat readiness. Maintaining this high level of combat readiness will necessitate frequent evaluations of individual and unit proficiency and the development of effective means to diagnose and remediate performance deficiencies.

The need for more frequent performance evaluations is not easily met. The Army's emphases on performance-oriented training and criterion-referenced testing already have increased the demand for training and evaluation uses of operational equipment and accompanying support resources requirements (e.g., fuel, ammunition, spare parts) during a time of inflation and budgetary constraints. More frequent performance-oriented individual and collective readiness evaluations would thus tend to strain an already tight situation.

A proposed solution to the problem of conducting more frequent evaluations in the face of tight resource constraints is to conduct these evaluations using training devices (e.g., simulators, mockups, etc.) instead of actual equipment (e.g., Finley, Gainer, & Muckler, 1974; Hopkins, 1975). Precedent already exists for employing simulators in this way. The commercial airlines and Federal Aviation Administration use flight simulators extensively in proficiency assessment. Also, the military is making greater use of training devices to evaluate individual and collective skills in other areas such as maintenance, antisubmarine warfare, and air defense artillery (Hawley & Dawdy, 1981a).

Improving the proficiency assessment capabilities of the Army is the idea behind the training Device Operational Readiness Assessment Capability (DORAC) concept (e.g., Finley, Strasels, Schendel, & Hawley, 1981). The concept suggests that training devices could and should be designed with the embedded capability for assessing operational readiness, and it would provide a cost-benefits method for training device proponents and developers to determine if and how to implement a DORAC in the design and utilization of any specific device (Hawley & Dawdy, 1981a, 1981b).

The research reported here was carried out to facilitate the implementation of a DORAC on Weaponer (WP), a rifle marksmanship trainer configured as an arcade type of device. More specifically, testing was conducted to determine if performance on WP can be used to predict soldier's M16A1 live-fire qualification, or record fire (RF), performance. RF is an annual requirement, and its conduct imposes time and cost burdens on commanders

who must billet and transport troops to remote RF facilities. This is a particular problem within U.S. Army, Europe (USAREUR) because of the scarcity of certified outdoor range facilities capable of satisfying RF requirements.

Method

Subjects. Sixty-nine Basic Rifle Marksmanship trainees at Fort Benning were divided into three equal groups. Assignment to groups was random with the constraint that high- and low-skill shooters be divided evenly across groups. Judgments about shooters' skills were based on progress charts maintained by company cadre.

Apparatus. WP is an M16A1 remedial marksmanship trainer designed to treat individual performance deficiencies (e.g., Schendel & Williams, in preparation). A non-restorable M16A1 rifle is equipped with a target sensor, and each target contains a light emitting diode (LED) which is sensed by the target sensor on the rifle. A predicted round impact point is determined by the LED-target sensor alignment. WP has a memory for recording up to 32 predicted shot impacts and a printer for providing a printout of all shots on selected targets. Recoil is simulated by the operation of a recoil rod which attaches to the barrel of the rifle. The sound of the rifle is transmitted through earphones.

WP has three targets: a scaled 25-m zeroing target, a scaled 100-m, "E-type" (kneeling) silhouette target, and a 250-m, "E-type" silhouette target. Targets are presented one at a time, but may be activated singly or in automated sequence by depressing the appropriate button(s) on WP's control panel or remote control box. The silhouette targets may be programmed to fall when hit by pressing the "kill" button. Exposure time may be varied from 2 to 30 sec for the 100-m target, 4 to 30 sec for the 250-m target, or set for continuous presentation. Firing pads used with WP provide the capability for the firer to fire from the foxhole supported (i.e., with sandbags) or prone positions.

A video display shows the shooter's aiming point which appears as a dot, or ball of light. The screen also displays the selected target and the location of hits (and misses). Two unique features of the video display are the "replay" and "each shot" controls. When activated, the replay feature shows the movement of the rifle 3 sec prior to firing, while the each shot feature displays the location of each shot fired in the order it was fired. The video display also includes: a numerical designator for the target on display, the number of hits on the target, the number of misses, late shots (fired after the target has dropped), the shot number, and the total number of shots fired (Spartanics, Ltd., 1976).

Procedure. Each soldier was allowed to fire, with feedback, one three-round shot group at the WP's 100-m target and one three-round shot group at the WP's 250-m target. Each soldier was then tested twice on a scenario, receiving 64 shots in all. Soldiers receiving the "8-24" scenario fired eight shots at the 100-m and 24 shots at the 250-m targets.

Soldiers receiving the "24-8" scenario fired the reverse pattern. Soldiers receiving the "Random" scenario fired 32 shots at a 50-50 random sequence of the two targets. The 100-m target always was presented for 2 sec, and the 250-m target always was presented for 4 sec. The intertarget interval was typically between 1 to 4 sec but occasionally was as long as 8 sec. Both targets fell when hit. Soldiers receiving the 8-24 scenario and the 24-8 scenario were tested using a single, randomized schedule of target presentations which was controlled remotely by the experimenter. For soldiers receiving the Random scenario, target presentations were under the internal control of the WP. All firing was done from the foxhole supported position. Soldiers rested briefly between the first and second set of 32 shots on a scenario and were provided feedback on the results of their performance only on completing testing.

RF occurred 24- to 48-hours after WP testing. Soldiers were assigned randomly to one of two firing lanes. The course-of-fire consisted of 14 "F-type" (prone) silhouettes seen at 50 (n = 5) and 100 (n = 9) m and 26 "E-type" silhouettes seen at 150 (n = 10), 200 (n = 8), 250 (n = 5), and 300 (n = 3) m. The first half of this course was shot from the foxhole supported position; the second half from the prone position. Targets fell when hit.

Results and Discussion

Three variables clearly influenced WP's ability to serve as a predictor of RF performance. These were scenario, soldiers' firing position at RF (foxhole supported position versus prone position), and first versus second set of 32 shots on WP.

Scenario. Overall performance on the Random, 8-24, and 24-8 scenarios correlated .66, .56, and .41 respectively with performance at RF. (In these and all other analyses that follow $r > .36$, $p < .05$; $r > .49$, $p < .01$). While performance on the Random scenario would appear the best predictor of RF performance, it probably is not. One soldier's data contributed disproportionately to the strength of this correlation. When this soldier's data were removed and the correlation recomputed, r dropped to .41.

In general, the 8-24 scenario produced the strongest correlations (first set of 32 shots, second set of 32, overall) with RF performance. The weaker correlations observed using the Random, and 24-8 scenarios probably resulted because these scenarios were too simple. Soldiers tested on the latter scenarios tended to show uniformly high scores on WP, but the usual wide variation in RF scores. For illustration, 52% of the soldiers tested using the 8-24 scenario hit over half the targets presented on WP. This compares to 78% of those tested using the Random scenario and 87% of those tested using the 24-8 scenario.

Firing position. Correlations of WP scores with the first half of RF were higher for all scenarios than correlations of WP with the second half of RF. This seems related to the change in position from foxhole

(first half RF) to prone (second half RF). The relationships for the 8-24 scenario soldiers were as shown in Table 1.

Table 1

Effect of Firing Position on 8-24 Scenario - RF Performance Correlations

	1st 32 <u>Shots</u>	2nd 32 <u>Shots</u>	<u>Overall</u>
RF Foxhole	.62	.74	.74
RF Prone	.15	.32	.26

Subsequent analyses of other data (i.e., Smith, Thompson, Evans, Osborne, Maxey, & Morey, 1980) suggest that the effect of firing position is not peculiar to our use of WP. These analyses revealed that practice RF is a significantly better predictor of RF performance when firing position is taken into account. Practice RF, like RF, includes 20 shots from a foxhole supported position and 20 shots from the prone position.

Set of 32 shots. Correlations for all scenarios were higher between performance during the second set of 32 shots on a scenario and RF performance than between the first set of 32 shots and RF. Again in the 8-24 scenario, performance on the second set of 32 shots correlated .60 with RF, while the first set of 32 shots and RF correlated only .44.

Conclusions and future directions. WP performance is a good predictor of RF performance and the prediction is enhanced when firing position and later shots are considered on the device. To confirm and extend these findings, further research will manipulate the strength of the relationship between WP performance and RF performance. In particular, soldiers tested on WP will (1) fire from both the foxhole and prone positions, and (2) receive varying amounts of training on the device prior to firing the 8-24 scenario.

A general conclusion is that implementation of a DORAC will require careful attention to testing and scoring methods. This may complicate the implementation process somewhat, but it should not be a particular problem when training device fidelity is high.

REFERENCES

- Finley, L. L., Gainer, C. A., & Muckler, F. A. Test and evaluation of a carrier air traffic control center team trainer as a performance qualification instrument: Phase I report. Canoga Park, CA: XYZYX Information Corp., 1974. (NTIS No. AD B003 213L).
- Finley, D. L., Strasel, H. C., Schendel, J. D., & Hawley, J. K. Training Device Operational Readiness Assessment Capability (DORAC): The concept and its implementation. In Proceedings of the 25th Annual Meeting of the Human Factors Society. Rochester, NY: Human Factors Society, 1981.
- Hawley, J. K., & Dawdy, E. D. Training Device Operational Readiness Assessment Capability (DORAC): Analysis and evaluation plan (ARI Technical Report). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences, February 1981 (a).
- Hawley, J. K., & Dawdy, E. D. Cost and Information Effectiveness Analysis (CIEA): A methodology for evaluating a training Device Operational Readiness Assessment Capability (DORAC) (ARI Technical Report). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences, February 1981 (b).
- Hopkins, C. O. How much should you pay for that box? In Proceedings of the 19th Annual Meeting of the Human Factors Society. Dallas, TX: Human Factors Society, 1975
- Schendel, J. D., & Williams, G. P. Rifle Marksmanship training and Weaponeer (ARI Research Report). Alexandria, VA: U.S. Army Research Institute for Behavioral and Social Sciences, in preparation.
- Smith, S., Thompson, T. J., Evans, K. L., Osborne, A. D., Maxey, J. L., & Morey, J. C. Effects of down-range feedback and the ARI zeroing target in rifle marksmanship training (Research Report 1251). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences, June 1980.
- Spartanics, Ltd. Weaponeer general information manual. Rolling Meadows, IL; Author, 1976

Development of Basic Skills Instruction for the Army:
Analysis and Verification of Basic Skill Requirements
For One Military Occupational Specialty

by

Eva Vaughan
Zita Glasgow

Applied Science Associates, Inc.

ABSTRACT

This paper describes the development of a Functional Basic Skills Education Program (FBSEP) for one Army MOS. The phases of the project include identification and verification of necessary basic skills and strategies, construction of a diagnostic test to measure them, and design of FBSEP instruction to teach them. A Gagne-type hierarchical analysis procedure was used to identify and analyze competencies necessary for success in the course. Competencies fell in traditional BSEP categories, e.g., reading, as well as some not-so-traditional ones, e.g., listening, equipment-related reasoning. Verification of prerequisites, a multi-stage process, has established the reliability of the analysis procedures and the relationship of identified prerequisites to AIT course success. The methods developed and used in this study seem applicable to the analysis of FBSEP needs and design of FBSEP instruction in other military occupations as well.

Development of Basic Skills Instruction for the Army:
Analysis and Verification of Basic Skill Requirements
For One Military Occupational Specialty

INTRODUCTION

The Army has long been concerned with the low level of basic skills of some recruits. Soldiers who are unable to read their technical and field manuals, and who lack basic quantitative and writing skills cannot fulfill the requirements of their jobs in the field. As a result, the Army has become one of the major teachers of remedial basic skills. Basic Skills Education Programs (BSEP) are now offered, at several levels of proficiency (BSEP I, II, etc.), at Army posts throughout the country and abroad.

However, with the exception of a few experimental programs, BSEP training differs little from reading, writing, and arithmetic instruction in elementary school. The skills taught are general skills; the methods include materials developed for teaching the skills to children; and the goals are generally expressed in terms of reaching a certain grade level, rather than acquiring particular skills.

There are several problems with this approach to basic skills training. First, the specific basic skills needed for carrying out military jobs often differ from those taught in BSEP. For example, while traditional reading instruction stresses reading-to-learn (reading to store information), Army reading usually involves reading-to-do (reading to perform an immediate task). Second, the materials and methods do not appeal to adults, especially adults who have been "turned off" by traditional instruction for some time. Third, the time allotted (usually a maximum of six weeks), is too short to raise general literacy skills to a level sufficient to deal with military materials.

Though general literacy skills are unlikely to improve significantly over a short period of instruction, significant changes can be produced by instruction more closely tailored to the soldier's job, that is, by functional basic skills training (FBSEP). For example, Thomas Sticht and his associates developed a functional reading skills program for several MOSs which, in six weeks, resulted in an average gain of 2.1 grade levels in job-related reading skills (Sticht, 1975).

For the past year, Applied Science Associates has been engaged in developing a FBSEP program for one particular MOS in the Signal Corps, the 31M "Multi-channel Communications Equipment Operator" MOS. Like many others, the 31M AIT course has been plagued by a high failure rate, at least partially due to low basic skill levels of many trainees. It is hoped that a basic skills program tailored to the 31M AIT course will produce students better able to cope with the specific reading, writing, arithmetic, etc. requirements of the course. The program is experimental in two ways. First, it attempts to identify not just traditional basic skills, but also strategies useful for learning from the modes of AIT instruction. Second, one of its goals is to develop a model for designing similar FBSEP programs for any MOS.

Before one can design instruction to teach the basic skills required by a particular course, one must determine what those basic skills are. Identification and verification of basic skills needed in the 31M course was the first

phase of the project. The second phase involves the development of a diagnostic test in order to identify students who need basic skills training prior to beginning the 31M course. And the third phase involves developing and validating units, modules, and lessons to teach the necessary skills. The first phase - identification and verification of basic skills prerequisites - is the subject of this paper.

IDENTIFICATION OF PREREQUISITES

Analysis of the Existing Course

Identification of prerequisites involved two stages: (1) analysis of the existing AIT course, and (2) analysis of prerequisites. The purpose of the course analysis was to identify the lowest-level skills actually taught in the course. The dividing line between these skills and still lower-level ones (prerequisites for the course), we called Baseline 1. Thus, the purpose of the course analysis was to identify Baseline 1.

Terminal objectives for each lesson and procedures for teaching them were obtained by examining lesson plans and by observing instruction being delivered. Gagne's hierarchical analysis method and classification system were then used to analyze each lesson (Gagne, 1977). The terminal objective of the lesson, behavior X, was classified in an appropriate category (association learning, discrimination learning, etc.). Next, enabling objectives (Y) were identified by asking, "What must a person be able to do in order to do X?" The procedure was repeated, iteratively, for each enabling objective, until the lowest level of behaviors actually taught in the course was reached.

Lesson analyses were submitted to course instructors for review and revised as necessary. We were now ready to move below Baseline 1.

Analysis of Prerequisites

Three methods were used to extend the hierarchical analysis below Baseline 1: (1) The lowest-level behaviors taught in the course were treated as terminal objectives (X), and enabling objectives (Y) were identified by the same methods as in the course analysis. (2) The materials used in the course (TMs, FMs, handouts, data forms, etc.) were analyzed to identify needed reading skills. (3) The instructional delivery system (lectures, demonstrations, guided practice) was analyzed to identify listening skills and learning strategies required to profit from these modes of instruction.

As we proceeded in the analysis, the distinction between basic skills and learning strategies often blurred. For example, should we call the ability to comprehend a paragraph describing job standards a reading skill or a strategy for extracting information from text? Therefore, we decided to use the more general term competency to describe any prerequisite, whether skill or strategy or both.

The competencies we identified fell into several subject-matter categories. Some of these are traditional subjects of basic-skill instruction: functional reading, functional arithmetic, and writing (filling out forms). Some are not traditionally taught in BSEP programs, or, at least, not stressed: equipment-

related reasoning, e.g., the kind needed in troubleshooting, technical information about radios and electronics; listening and note-taking during lectures and demonstrations. The latter seemed to us to be particularly important in courses like the 3IM, which much of the instruction is oral and all of it is instructor-paced. Yet, such competencies have received little attention in Army BSEP, or, for that matter, in basic skill training at any place or any level.

VERIFICATION OF PREREQUISITES

Before instruction to teach competencies can be designed, it is necessary to verify that the right competencies have been identified. This involves answering a number of questions. (1) How reliable was our analysis process? That is, do two or more analysts, starting from the same point and working independently, identify the same competencies, in the same hierarchical relationships? (2) Can we develop objective, reliable, and valid measures of the competencies? (3) Are the hierarchical relationships correct? That is, are the lower-level skills in each hierarchy truly prerequisite to the higher-level ones? (4) Are the competencies related to performance in the 3IM course? (5) Does teaching the competencies to students lacking them increase their probability of success in the course?

As the questions imply, verification is a multi-stage process. The first two have been completed; the third stage data collection is complete and is being analyzed; the fourth and last is not possible until FBSEP has been designed and implemented.

Stage I Verification

Stage I Verification, conducted early in the analysis, was designed to verify the reliability of the identification/analysis process. Three analysts, working independently, analyzed a single high-level course prerequisite into its components. Both percentage agreement on competencies identified and correlations among hierarchical order were high (average of 75% and .91, respectively). We concluded that competencies can be reliably identified, using our method.

Stage II Verification

The purpose of Stage II Verification was to try out some preliminary measures of competencies to obtain student perceptions of course needs, to find out the level of functioning of 3IM students on some of the competencies, and to examine relationships between competencies and course performance. A sample of 27 good and 35 poor¹ students were interviewed, and were asked to complete a series of exercises designed to measure selected competencies. Some of the results are described below.

Given a brief lecture and told to take notes, good students tended to take notes which included more information and were better organized than those of poor students. Moreover, both good and poor students rated the ability to listen and take good notes as extremely important for success in the course, thus corroborating our hunch about the importance of these kinds of skills.

¹ Course instructors selected good and poor students, based on formal and informal performance measures.

Technical information concerning radio and electronics varied from very low to very high in both good and poor students, suggesting that such knowledge is not important for success. This was confirmed by low ratings of importance from both instructors and students. As the course is now taught, technical information is not necessary to learn to operate electronically complex radio equipment.

In reading, both good and poor students possessed adequate decoding skills. Comprehension, however, varied widely. Differences between good and poor students were especially evident in interpreting procedural directions, which are often presented, in TMs, in syntactically complex forms. Other reading skills which varied among students included word knowledge and locating information in tables of contents, and in troubleshooting and maintenance tables.

Student interviews suggested a tendency for good students, more often than poor, to mention the use of visual imagery as a strategy for encoding information in lectures, demonstrations, and films. Given the nature of the skills learned, the modes of instruction, and the types of students in the 3IM course, imagery seems to be a particularly relevant - and potentially teachable - strategy.

The exercises we had developed seemed effective as measures of the competencies, and served as prototypes for development of test items in the next stage.

Thus, Stage II Verification provided information verifying the relevance of certain competencies (e.g., note-taking, listening, reading comprehension), and the irrelevance of others (e.g., technical information). In addition, it contributed to analysis, which was continuing at the same time, by filling in some holes in the hierarchies.

Stage II Verification and the subsequent completion of the analysis of prerequisites resulted in an organized list of partially verified competencies in the following areas:

1. Equipment-Related Reasoning Skills
2. Quantitative Skills
3. Reading Skills, consisting of:
 - Reading Connected Discourse
 - Reading Labels to Identify Equipment Components
 - Reading to Locate Information
 - Reading a Cabling Diagram
4. Writing Skills (other than note-taking)
5. Auding (listening) Skills
6. Note-Taking Skills

Stage III Verification

We are currently completing Stage III Verification. Its purposes include: (1) tryout of multiple-choice questions designed to measure the competencies; (2) validation of hierarchical relationships between pairs of competencies; (3) determination of relationships between clusters of competencies and objective measures of course performance of 3IM students. For these purposes, at least six items were written to measure each competency, and the resulting test was administered to 163 students beginning the course and 93 others completing it.

Content validity was built into the test by using the organized list of prerequisites as a table of test specifications, and by selecting the content from actual 3IM course materials whenever possible. Item analysis statistics (item difficulties and correlations with total competency scores) and split-half reliability coefficients have been computed for various competencies and competency clusters. The results of item analysis include:

- The generally high correlations of item scores with total competency scores suggest that our competency measures are indeed measuring unitary, homogeneous competencies.
- Item difficulties vary greatly from one competency to another. Some competencies, especially high-level reading comprehension ones) are extremely difficult even for students completing the course. This suggests that such competencies are not truly necessary for course success, probably because the structure of the course makes it possible for students to learn via alternate modes. For example, poor readers may succeed by using listening skills to acquire the same information which good readers acquire from the TM.

Split-half reliability coefficients for competency clusters varied from .363 to .945. As expected, the reliability was closely related to number of items in the cluster. For clusters with 12 or more items, coefficients were .60 or more with one exception. Since the test used in Stage III Verification was a tryout version of the diagnostic test to be constructed in the next phase of the project, steps will be taken to ensure the reliability of measurement of all competencies verified as important for the course.

For validating hierarchical relationships between pairs of competencies, we are computing Walbesser's (1968) consistency ratio and adequacy ratio, and Capie and Jones' (1971) necessity ratio for each pair. The results will help us to sequence instruction to teach groups of hierarchically related competencies.

Using data from students who have completed the course, we are correlating cluster scores with 3IM weekly performance test scores and with end-of-course test scores. The results will tell us which kinds of competencies are and which are not related to course success, and will help in setting priorities for skills to be included in FBSEP.

SUMMARY AND CONCLUSIONS

Stage III Verification will result in a final list of competencies recommended for training. In addition, the test analyses are feeding into diagnostic test construction, while the hierarchical validation and relationships with course performance are contributing to FBSEP design. Thus, each phase of the project depends upon the previous one and leads to the next. The final verification of our entire effort will come when we deliver FBSEP training to students who need it and determine whether their course performance improves as a result. We believe that the methods we are using constitute a viable means of designing successful FBSEP training programs for adults in the military, and possibly, in the civilian sector as well.

REFERENCES

- Capie, W., & Jones, H. An assessment of hierarchy validation techniques. Journal of Research in Science Teaching, 1971, 8, 137-147.
- Gagne, R. M. The conditions of learning (3rd ed.). New York: Holt, Rinehart, and Winston, 1977.
- Sticht, T. G. A program of Army functional job reading training: Development, implementation, and delivery system (HumRRO FR-WD-CA-75-7). Alexandria, VA: Human Resources Research Organization, June 1975.
- Walbesser, H. H. A hierarchically based test battery for assessing scientific inquiry. Paper read at the American Educational Research Association Meeting, 1968.

EQUIVALENCY OF SIMULATOR EXERCISES TO REAL WORLD EXPERIENCE IN MEETING IMCO STANDARDS

Kent E. Williams and Phyllis J. Kayten
National Maritime Research Center (CAORF)

Abstract

Recently the IMCO International Convention on Standards of Training, Certification and Watchkeeping for Seafarers agreed to require one full year of at-sea time for certification and licensing of maritime academy graduates to be third mate merchant ship deck officers. This project researched the questions of what types of Navigational Watchstanding tasks can be trained on a simulator that might qualify for equivalency to real world experience, and how many hours of simulator training are necessary to promote the acquisition of skills such that a simulator-trained subject's performance level would equal that acquired by a typical merchant marine cadet at sea. All research and training was performed at the Computer Aided Operations Research Facility (CAORF).

Fifteen first-class cadets at the U.S. Merchant Marine Academy who had had a nominal year at sea provided the criterion performance levels for a 50 minute simulated watchstanding exercise. Twelve fourth-class cadets who had no real sea experience were given a combination of classroom and simulator training aimed at teaching watchstanding skills equal to the level supposedly reached by the cadet during the required sea duty. Six of the fourth-class cadets were given 24 hours and six were given 48 hours of simulator training exercises. Expert masters viewed video tape recordings of watchstanding performance of the first-class cadets and of the fourth-class cadets before and after simulator training. Simulator training resulted in significant improvement in watchstanding performance of both 24 and 48 hour simulator trained groups. Fourth-classmen performed significantly worse than first-classmen before training and were in most performance categories not significantly different from the sea experienced first-classmen after training. Some skills appear to be more effectively trained on the simulator than others.

Introduction

The training of watch officer personnel has long been a major practice of all maritime nations as evidenced by their respective professional academies. It is no surprise then that the application of simulators as training devices would reach into those environments to promote the development of highly skilled professional watch officers. However, its application at the level of the maritime academies has been somewhat slow due to the existence of school training ships and apprentice ship assignments on operating merchant vessels. A major international convention and recommendations there from, as well as significant increases in fuel costs and significant advances in training technology have spurred on the utilization of shiphandling simulators as a viable cost effective and in some cases, a superior approach to the training of merchant academy cadets. The international events which have focused the application of simulation for cadet training resulted from the IMCO International Convention on Standards of Training, Certification and

Watchkeeping for Seafarers. This convention held in London, 14 June through 7 July, 1978, presented the agreement of 72 nations concerning the establishment of basic requirements for training for mates, officers and crews of seagoing merchant ships. Of critical importance to officer training programs was Regulation II/4 "Mandatory Minimum Requirements for Certification of Officers in Charge of a Navigational Watch on Ships of 200 Gross Registered Tons or More". Section 2(c) of this regulation introduces a requirement for one "actual" year of seagoing service for original certification and the year must include six months of bridge watchkeeping duties under the supervision of a qualified officer. This requirement poses a real problem for a number of existing officer training programs which do not include this actual year of time at sea. Among these are the programs at the maritime academies. Although the U.S. Merchant Marine Academy has a "nominal" year of sea time in its course of study, the "actual" time at sea is approximately ten months. The state academies on the other hand primarily depend upon the training ship cruise system and may have a total sea time considerably less than the required year.

Fortunately, the convention addresses itself to "Equivalents" in Article IX; Section (1) of this Article suggests that alternate training methods might be accepted if they can be shown to have effects equivalent to regulation requirements.

This provision therefore, opened the arena for the use of simulation to supply partial credit for at-sea duty time and consequently, allowed for a cost-effective alternative to at-sea time especially for those institutions encumbered by the high costs of operating school ships. The questions, however, still remain as to what types of tasks can be trained on a simulator that would qualify for equivalency to real-world experience and how many hours of simulator training are necessary to promote the acquisition of skills on the simulator such that performance would equal that acquired at-sea.

Another fortunate event was taking place in parallel to the activities of IMCO. This event was a joint program undertaken both by MarAd and the USCG in anticipation of the role which simulators would play in the training and certification of watch officers. These agencies had the foresight to plan a difficult program to determine the types of tasks and shiphandling objectives which could best be acquired on simulators for all levels of watch officers as well as to determine the simulator hardware and software requirements necessary to provide effective training of individuals with various input characteristics (i.e., masters, chief mates, second mates third mates and cadets). As a result of the ongoing program, numerous investigations have been published (Hammel and Williams, et. al, 1978; Friedberg, 1979; Fiore, 1979; Gynther et. al, 1979; Williams, 1980 and Gardenier, 1979; see also the Fourth Annual CAORF Symposium 1980).

As a logical consequence in the flow of events and programs addressing the role of simulation in the training and licensing process, the to be presented project was proposed. Again, the major question to be addressed by this investigation was the equivalency of simulator training to real-world experience in the acquisition of specific watchstanding tasks. The watchstanding tasks in question are those which must be experienced by the cadet and logged in their cadet training file as a function of their at-sea apprentice assignment on operating merchant ships. The cadets in question here are restricted to only those individuals actively participating in the Federal academy program at the U.S. Merchant Marine Academy, Kings Point, New York.

Method

Objectives. The purpose of this research was to examine the equivalence of varied amounts of simulator experience to simulator-applicable aspects of the "one-year" sea time experience required of marine academy graduates. First classmen of the Merchant Marine Academy at Kings Point, New York with nine to twelve month's at-sea experience were compared with fourth-classmen, who had completed one year of classroom study but had no at-sea assignment. The fourth-classmen and first-classmen were compared on basic performance measures reflecting knowledge of three training objectives: fundamental shiphandling, (i.e. ship characteristics), rules of the road, and watchstanding procedures.

Subjects

Three groups of subjects were used: (1) fifteen first-classmen who had completed nine to twelve months at-sea duty and who had successfully completed third mate licensing, (2) six fourth-classmen with no at-sea duty who completed a course including twenty-four hours of simulator time (four hours hands-on experience per individual) with six hours of classroom lecture, and (3) six fourth-classmen with no at-sea duty who completed a course including forty-eight hours simulator time (eight hours hands-on experience per individual) with twelve hours of classroom lectures. Since the question of interest was equivalence of simulator training to at-sea experience, selected measures of performance on a simulator diagnostic scenario by the experienced first-classmen were used as baseline performance data. The same diagnostic was administered to the two groups of fourth-classmen before simulator training and after their twenty-four to forty-eight hour simulator training.

Experimental Variables

The one variable of interest, then, was the amount of simulator training system experience needed to approach or approximate baseline performance of experienced first-class baseline data. Performance differences on the diagnostic between first-classmen and fourth-class groups after the fourth-classmen received simulator training provided a measure of simulator at-sea equivalence, and the difference between twenty-four hour and forty-eight hour training groups provided a measure of gain with increased amounts of simulator experience.

Description of Training Program

The training program for this study was designed to provide fourth-class cadets with accelerated experience in specific aspects of fundamental shiphandling (i.e., ship characteristics), Rules of the Road, and bridge procedures in an attempt to determine equivalence to at-sea experience required before licensing. Following in pre-test for all cadets, the two groups of six fourth-classmen participated in hands-on simulator exercises and classroom lecture/discussion of events and topics relative to the three training objectives mentioned above. The design of the experiment along with the flow of events is presented in Table 1.

Pre-Test/Post-Test Diagnostic Specifics

The pre/post-test diagnostic scenario was designed to elicit watchstanding tasks associated with skill and knowledge requirements assumed to have been learned at-sea during the midshipman apprentice ship assignment on operating merchant vessels. A considerable number of tasks and procedures must be integrated by the watchstander in order to satisfactorily make the vessel passage. The scenario evolved within the CAORF New York Data Base. (See Figure 1)

The cadets entered the bridge of the simulated ship, which was in motion and on its way to arrival at Ambrose via the Barnegat Traffic Lane. Each cadet had been given "standing" and "arrival" orders instructing him/her to:

1. Arrive abeam buoy "BA" at 0840 and adjust speed accordingly.
2. Switch to hand steering 7.0 miles south of "BA" buoy.
3. Notify pilots at buoy "BA" of 0930 arrival and adjust speed to make ETA at pilot station.
4. Engine room should also be notified of arrival times.
5. Call master five minutes before abeam of buoy "BA".
6. Test steering gear and engines astern prior to "BA" buoy.
7. Notify master if in doubt at any time.

During the test scenario, the cadet was expected to take over the watch from a confederate experimenter-watchstander, who offered the same information to all cadets unless the cadet requested additional information. The watch lasted from a simulated 0750 to 0840, at which time the confederate "master" arrived on the bridge to be briefed. Cadets were evaluated on their performance of required tasks. Their expected performance includes the following chronology:

0750 Watch Transfer — relieving mate should verify the following information:

- Course/compasses (visual)
- Vessel's position (through fixes)
- Speed (distance run/time from chart)
- Active targets (visually/radar, radar plotting)
- Night/special orders (read and sign)
- Familiarization with watch period specifics (future course or speed changes, currents, obstructions, navigation aids, logbook entries, etc.)

0800-0840 — Steady State Watch

- Monitor vessel's position along plotted track
- Monitor traffic and assess (course changes, speed changes, communication)
- Observance of standing orders (speed adjustments for arrival times, communication with pilots, engine room, master, etc.)
- Manage watch (port preparation; pilot ladders, etc., interactions with helmsman, lookout)

0840-0900 — Arrival State Watch

- Brief master on watch status (active traffic, communication, vessel course and speed, etc.)

During the steady state watch, the cadet is presented with traffic, both meeting and passing, and must abide by Rules of the Road and prudent procedures to avoid close encounters. Also during the steady state watch the steering engine is made to fail, and the cadet must take proper measures in order to facilitate restoring of steering.

Measurement of Performance

On completion of all tests and training, six expert mariners from the New York chapter of the American Master Mariners Association served as judges of performance. Masters were chosen because of their significant personal interest in judging and determining performance levels of prospective officers of these masters' ships. Their objective was to view the video tape recordings of the performance on the simulated watchstanding scenario of the six first-class midshipmen who had completed the nominal year at sea, and the twelve fourth-class midshipmen before and after simulator training. There were 30 video tapes in all, consisting of recordings of pre-tests and post-tests for the 12 fourth-classmen (24 tests) and pre-tests only for the six first-classmen. Each tape lasted the length of the diagnostic scenario, approximately 50 minutes, totaling almost 30 hours of viewing time.

Presentation of the tapes was "blind". The raters were not informed of the class status of the cadet being viewed, nor were they told whether they were viewing pre- or post-training performance. The raters were aware that the tapes would contain repeat performances by some of the midshipmen. However, the picture resolution was not high enough to allow recognition of personal characteristics and the cameras were more frequently directed at instruments and displays than on the cadets. None of the raters admitted to recognizing any cadet from previously viewed tapes.

Scoring of Forms

The form was scored in the following manner:

The evaluation form was broken down into four parts, corresponding to four major portions of the watchstanding scenario. The four sections are clearly delineated in the evaluation form: (1) Watch Transfer, (2) Steering Failure, (3) Steady State Watch (hereafter referred to as Watchstanding) and (4) Briefing Master.

Scores for each section were computed separately. Scores for the Watch Transfer, Steering Failure and Briefing Master were all computed in the following manner: The total number of boxes (activities) checked either "yes" or "no" comprised the total valid answers scored. The total score for these three performance categories was equal to the percentage of boxes checked "yes" out of all valid answers scored. Thus, if a box was unscored, it was treated as non-applicable rather than as a necessary action which was not taken. Cadets, then, were not penalized if their behavior was appropriate, even if they did not perform a particular task listed in the rating form.

The Watchstanding (Steady State Watch) category was scored by computing the average rating of all five point scales actually scored by the rater. If a question was not answered, i.e. no number was circled, this question was not included in the average.

Results

Performance Differences at Time of Pre-Test

Table 2 depicts the pre-test comparisons of experimental groups. Differences between the first-class, fourth-class 24 hour and fourth-class 48 hour training groups were analyzed separately for the four performance categories: (1) Watch Transfer, (2) Steering Failure, (3) Briefing Master, and (4) Watchstanding.

Because of the extremely small amount of valid data within some groups, statistical significance was very difficult to achieve. The lack of valid data was a product of two factors. The first factor was the unavoidable small number of subjects in each group. The second factor is somewhat more significant: the masters left many questions blank because they felt they were not applicable or not clear enough. This reaction led to the redesigning of the evaluation format, for future research projects. Some differences, however, were significant, and are reported here.

With respect to pre-test group differences, the three experimental groups were not statistically different for the performance categories of Steering Failure and Briefing the Master. In the Briefing Master category, it appears that the first-classmen were much better than fourth-classmen in performance, but the numbers of subjects in the fourth-class group cells were too small to allow for statistical significance.

In the Watchstanding category, first-class cadets were significantly better than the two fourth-class groups, which did not differ from each other. This difference is what would be expected at pre-test. For the Watch Transfer category, the fourth-class 24 hour training group scored significantly lower than the first-class and the fourth-class 48 hour group. No explanation can be given for the surprisingly high average score of the 24 hour fourth-class group. Although their average score was relatively lower than the first class group, statistically this difference was not significant.

Performance Differences at Time of Post-Test

Table 3 depicts the performance differences between the first-class group and the two fourth-class groups after training. No significant differences were found between any group on any performance measure. On two measures, Steering Failure and Watchstanding, the two fourth-class groups had higher average scores than the first-class group although these differences were not significant. It should be noted, however, that the fourth-class groups were significantly worse than the first-class group in the Watchstanding category before training (see Table 2). After training this difference disappeared.

Effectiveness of Simulator Training - Main Effect and Time by Training Interaction

Tables 4 and 5 depict the main effect of simulator training from pre- to post-test and the differences in training effect due to amount of training, respectively. A statistically significant increase in performance level from pre- to post-test (Table 4) was found across fourth-class groups for the Steering Failure and Watchstanding performance measures. The increased performance level from pre- to post-test for the Watch Transfer performance category was not statistically significant. The Briefing Master category showed no significant change, although this finding is due to the lack of valid data. From the data available, the pre- to post-test differences appear to be high based on visual inspection, and would be expected to be statistically significant with more valid data.

The differences depicted in Table 5 between the 24 and 48 hour training groups are rather hard to interpret. There were no significant differences between the one week 24 hour and 48 hour groups in pre- to post-test change for any performance category except for Watch Transfer. In the Watch Transfer category, the 24 hour group showed a change from pre- to post-test statistically different from the 48 hour group. In fact the 48 hour group showed a performance decrement while the 24 hour group showed a performance increment. Although it is difficult to speculate on these differences given the small number of subjects in each group, it is possible that the behaviors rated in the Watch Transfer section were not stressed in later sessions, and the cadets in this group may have felt that they did not need to ask the mate on watch for information they had learned how to gather themselves. If this was the case, instructors should be warned to stress this important step in watchstanding throughout the training session.

Discussion

Simulator training was found to have a significant effect on watchstanding performance. Because of the relatively small range of scores given by masters, the difference in training effect from 24 to 48 hours is difficult to determine. It appears from this evaluation that a greater range of training hours must be tested in order to evaluate the differential training effect of varied hours of simulator experience. The decrease in performance from 24 to 48 hours in Watch Transfer (see Table 5) cannot be explained at the present time. It appears that the masters used a different set of priorities when evaluating watchstanding, and in particular, Watch Transfer performance than was anticipated in the development of the rating form. This difference in rating (and training) priorities was discussed in depth during the evaluation procedures. The results was a restructuring of the evaluation form, and suggestions for areas of emphasis in training. Use of the new evaluation form in future training research will most likely decrease any ambiguity in analysis of results.

Analysis based on the preliminary form, however, yielded consistent ratings from the masters, and clearly indicated an increase in performance levels after simulator training. These results further suggest some measure of equivalency between performance gain due to at-sea experience and that due to simulator experience for specific deck watchstanding tasks.

References

- Fiore, A. and Lentz D.D. Midshipmen/Cadet Training Experiment Third Annual CAORF Symposium, Kings Point, New York, October, 1979.
- Friedberg, A.W. Shiphandling Simulators For Cadet Training Third Annual CORF Symposium, Kings Point, New York, October 1979
- Gardenier, J.S. Where Are We Headed With Bridge Simulators? Third Annual CAORF Symposium, Kings Point, New York, October 1979.
- Gynther, J., Gaffney, M. and Gilder, M. CAORF's Master/Chief Mate Simulator Training Experiment Third Annual CAORF Symposium, Kings Point, New York, October, 1979
- Hammell, T.J., Williams, K.E., Evans, W. and Grasso, J. Role of Simulators in the Mariner Training and Licensing Process, Phase I. National Maritime Research Center, U.S. Maritime Administration, Contract No. 3209-000, November, 1978.
- Williams, K.E. Validation of Valdez Operational Exercises Training Employing a Backward Transfer Methodology Fourth Annual CAORF Symposium, Kings Point, New York, September 1980

Figure 1.

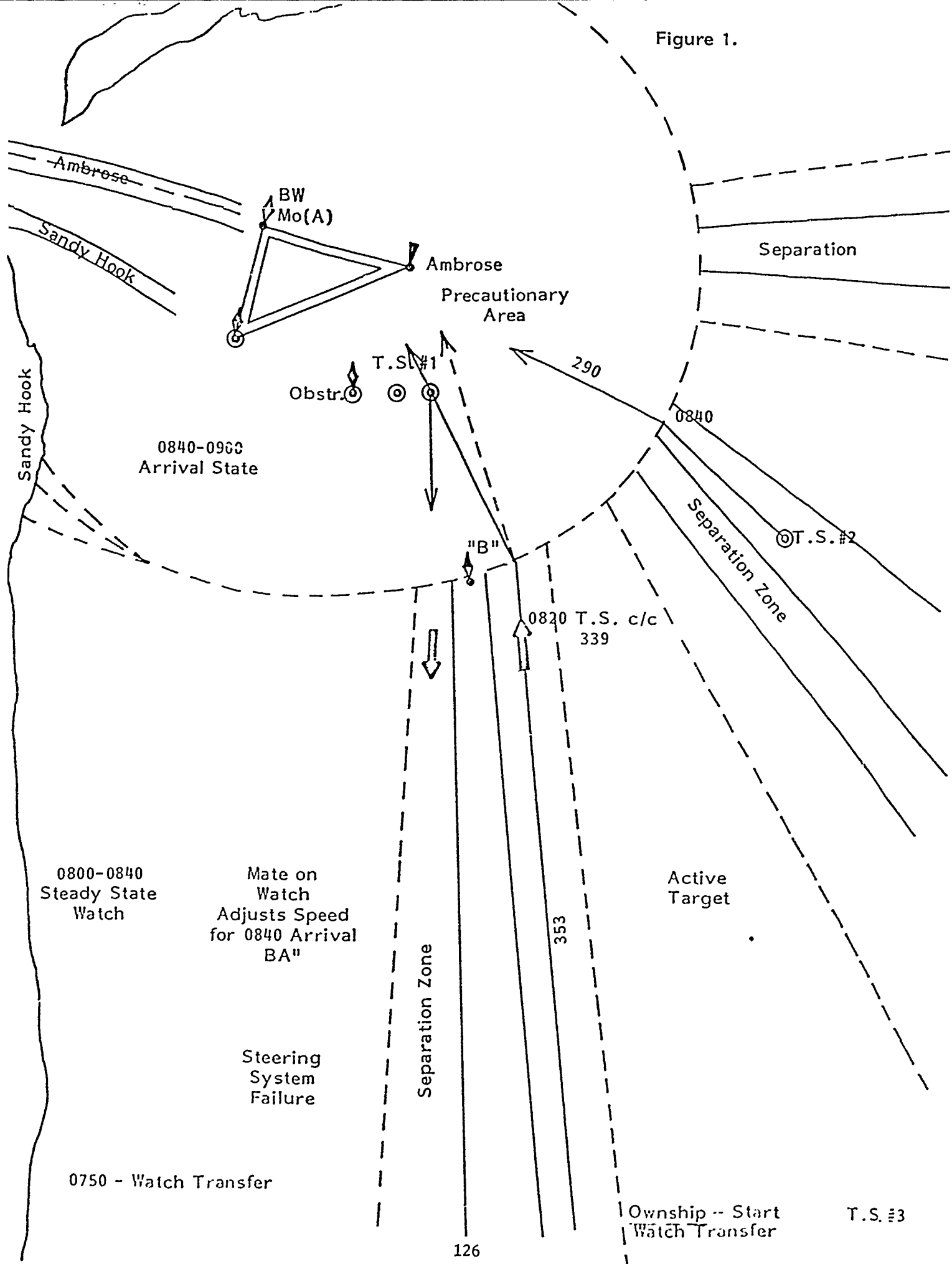


TABLE 1
CADET EQUIVALENCY TRAINING RESEARCH
DESIGN SCHEDULE

I. ORIENTATION	II. PRETEST 'A'	III. TRAINING PROGRAM	IV. POSTTEST 'B'
<u>First Classmen</u> Familiarization with bridge equipment	X		
<u>Fourth Classmen Group 1</u> Familiarization with bridge equipment	X	24 hours hands-on four hours per trainee 6 hours classroom	X
<u>Fourth Classmen Group 2</u> Familiarization with bridge equipment	X	48 hours hands-on eight hours per trainee 12 hours classroom	X

Comparisons will be made between Fourth Classmen and First Classmen on Diagnostic A, and between Fourth Classmen performance on Diagnostic B with First Classmen on Diagnostic A. The latter comparison will yield the degree of equivalency between simulator and at sea trained cadets. Comparisons between Fourth Classmen groups on change scores (B-A) will be made to determine effect of additional simulator training on performance.

X - Means Pretest or Posttest will be administered.

TABLE 2. ANOVA OF PRETEST COMPARISONS OF FIRST CLASS, FOURTH CLASS ONE WEEK SIMULATOR TRAINED AND FOURTH CLASS TWO WEEK SIMULATOR TRAINED GROUPS AS RATED BY EXPERT MARINERS

Performance Measure	First Class Group		One Week Fourth Class Training Group		Two Week Fourth Class Training Group		F Ratio
	Mean	SD	N	Mean	SD	N	
1. Watch Transfer - Percent Behaviors Checked Satisfactory	86.8	18.3	5	23.7	29.7	3	6.4*
2. Steering Failure - Percent Behaviors Checked Satisfactory	61.2	34.7	4	62.0	22.8	5	.3
3. Briefing Master - Percent Behaviors Checked Satisfactory	72.6	43.6	5	0.0	0	2	3.2
4. Watch Standing - Subjective Evaluation Total	2.06	.37	5	1.35	.2	5	9.77**

*Significant difference with probability = .02. One Week Training Group significantly lower than other groups by post-hoc comparison.

**Significant difference, $p = .003$. First Class significantly higher than Fourth Class Groups by post-hoc comparison.

TABLE 3. ANOVA OF POSTTEST COMPARISONS OF FIRST CLASS GROUP, FOURTH CLASS ONE WEEK SIMULATOR TRAINED AND FOURTH CLASS TWO WEEK SIMULATOR TRAINED GROUPS AS RATED BY EXPERT MARINERS

Performance Measure	First Class Group (Pretest Scores)		One Week Fourth Class Training Group		Two Week Fourth Class Training Group		F Ratio
	Mean	SD	Mean	SD	Mean	SD	
1. Watch Transfer - Percent Behaviors Checked Satisfactory	86.8	18.3	63.0	10.4	50.0	41.2	2.12
2. Steering Failure - Percent Behaviors Checked Satisfactory	61.3	34.7	75.0	10.0	85.0	19.1	1.02
3. Briefing Master - Percent Behaviors Checked Satisfactory	72.6	43.6	62.5	47.9	65.7	47.2	.05
4. Watch Standing - Subjective Evaluation Total	2.1	.37	2.6	.68	2.4	0.5	1.4

NOTE: No significant differences between Groups were found after training.

TABLE 4. ANOVA OF PRE AND POSTTEST DIFFERENCES-COLLAPSED OVER FOURTH CLASS TRAINING GROUPS AS RATED BY EXPERT MARINERS (TRAINING MAIN EFFECT)

Performance Measure	Pretest Score			Posttest Score			F Ratio
	Mean	SD	N	Mean	SD	N	
1. Watch Transeer - Percent Behaviors Checked Satisfactory	44.8	40.3	5	54.2	35.6	5	0.3
2. Steering Failure - Percent Behaviors Checked Satisfactory	56.25	19.9	8	80.0	15.1	8	7.2*
3. Briefing Master - Percent Behaviors Checked Satisfactory	0	0	2	81.5	26.2	2	19.4
4. Watch Stranding - Subjective Evaluation Total	1.4	.19	8	2.5	.56	8	40.6*

*Significant difference with $p \leq .05$.

TABLE 5. ANOVA OF COMPARISON OF ONE AND TWO-WEEK SIMULATOR TRAINED FOURTH-CLASS GROUPS IN THE CHANGE BETWEEN PRE AND POSTTEST PERFORMANCE AS RATED BY EXPERT MARINERS (TRAINING X TIME INTERACTION EFFECT)

Performance Measure	One-Week Group			Two-Week Group			F Ratio
	Mean	SD	N	Mean	SD	N	
1. Watch Transfer - Percent Behaviors Checked Satisfactory	50	9.9	2	-17.7**	22.5	3	14.8*
2. Steering Failure - Percent Behaviors Checked Satisfactory	12.5	27.5	4	35.0	19.1	4	1.8
3. Briefing Master - Percent Behaviors Checked Satisfactory	100	0	1	63	0	1	N/A - (n of 1 in each group)
4. Watch Standing - Subjective Evaluation Total	1.3	.6	4	0.9	.4	4	.9

*Significant difference with $p \leq .05$.

**A negative score indicates decrease in behaviors checked satisfactory by rater.

TRAINING PROGRAM EVALUATION:
AN APPLICATION OF DISCREPANCY EVALUATION
TO MILITARY TRAINING

Donald M. Kristiansen
US Army Research Institute - Fort Knox Field Unit

The primary mission of the Army is to wage war. When the Army is not waging war, its primary mission becomes training to wage war. Leaving aside all the problems with constraints to training, training distractors, etc., training is taking place throughout the Army every day; institutional training programs, transition training programs, unit training programs, school training, individual training, collective training, and special training (Army Correspondence Course Program, Training Extension Course Program, etc.).

This training has been developed in response to certain needs and, hence, has goals. These goals are usually expressed in terms of objectives. Training programs are developed to achieve the objectives, tests are given to assess soldier proficiency, and the process usually stops there. The Army has a penchant for developing programs, directing their implementation, and then making three rather large assumptions: 1) that the program design will, in fact, realize the program objectives, 2) that the program has, in fact, been implemented in accordance with program design, and 3) that the products of the program (trained soldiers) do, in fact, meet the original goal. Too often, these assumptions are not warranted but no one knows that, because there are few normal ongoing training program evaluation systems in the Army (GAO, 1981).

Training institutions (centers and schools) do not routinely examine the training process (Hall et al., 1976). Directorates of Evaluation are usually charged to ensure that the right things are being taught to the right standards. But, there is some question, usually, as to what should be taught in the first place (since centers often doubt the validity of their front-end analyses), and either standards are not firm or they are excused as being inappropriate for this stage of training, unachievable given current resources, or given scant attention. Transition training is usually examined in terms of the hardware operations rather than the training process, the question being, "Are we training these soldiers to correctly operate that piece of equipment?" Units seldom examine in-unit training from the process standpoint, the focus being on the product, terminal soldier proficiencies.

For those of you who have closely examined training, as a naturalist would, unobtrusively observing what goes on, you know what the results of this lack of attention to training process has been. For those of you who have not been able to do that, let it suffice to say that in many instances, the training process has not been tightly managed and shows the direct effects of having been free to become instructor specific. Proficiency is determined by who trained whom, today.

Where this situation exists, it can be turned around if the chain of command would begin to pay attention to the training process. Most commanders would like to but they do not have materials and guidance, written for them,

that tells them, in detail, how to do this. ARI at Fort Knox has addressed this need and has produced and piloted a system for use by commanders who want to attend to the training process, at the school, center, or unit level.

The ARI system, called Training Program Evaluation (TPE), is a variant of a system developed by Malcolm Provus and his associates in the Pittsburgh school system in the early 1970's in response to needs to tell the federal government how effectively they were spending educational grant money (Provus, 1971). TPE is also similar to systems reported by Stake (1967) and Sanders and Cunningham (1973). Provus and his associates, at the beginning of their studies, thought they knew what was going on in these programs. Then they tried to describe, in an objective fashion, what was happening (tried to be data based) and found out that they did not really know what was happening. And they did not have ready procedures for finding out. On investigation, one federally funded remedial training program was described as 150 different programs, each one unique to a single teacher. To solve the problem of evaluating the effectiveness of educational programs, Provus and his co-workers developed a system called Discrepancy Evaluation. This system worked well for them, and the application we have come up with for use in the Army appears to work well too, at least in the applications conducted so far.

I want to describe for you, Discrepancy Evaluation, and our program, TPE, to show how they are similar, and describe the TPE applications that have led to an ever widening interest in the Army training community.

The first problem Provus found was that the goals were not specified clearly enough for everyone concerned to get a firm grip on what the program was supposed to accomplish. Goals were expressed in terms that defied evaluation. The Army does better than this, but not always. Goals, in the Army, are expressed as training objectives. Many training objectives are not measurable. We still familiarize people, we still orient them, we still require performance in accordance with section so-and-so of such-and-such a document which is, in fact, thirty or forty pages long and contains no clear description of the task.

The second problem Provus found was that, in many cases, the educational program, as designed, would just not realize the objective. The design was poor. Even if the program was faithfully carried out, the objective would probably not be achieved. In the Army, we have a recognized process that training design is supposed to go through to meet this problem; ISD or some variant. In fact, little training is designed by people expert in this process. The major problems they have are covered by ISD phases III.1 (Specify Learning Events/Activities), III.2 (Specify Instruction Management Plan/Delivery System), and III.5 (Validate Instruction) (Schulz et al., 1979). What usually happens is that subject matter experts, some NCOs, are tasked to develop a training package. They do the best they can but they are being tasked to do a job they have not been trained to do. Then, the Army seldom follows up to determine if the training program they have designed will, in fact, if faithfully implemented, realize the training objectives.

Next, Provus found that programs were not being implemented as designed. Each teacher was left to his or her own devices to interpret the objectives,

including the tasks and standards, and were seldom conversant with the design. Often, implementation did not differ from design so much because of constraints as because the teachers did not want to use the design; they liked theirs better. The Army does not often check to see if training design is, in fact, being faithfully implemented. Not only are standards given little attention but tasks are not trained, other tasks (not specified) are substituted, demonstrations are not given, and practice is not only not given, in some instances it is not allowed. The overriding notion seems to be that training is what the instructor does, rather than what the soldier does.

Provus decided that product measurement did not make much sense if you did not know what process had produced the product. If the product does not meet the requirements of the training objectives, you have no information on which to base modifications so that you do get the product you want next time. If the product is all right, you have no guarantee that it will remain all right unless you can specify that the process is stable. We do a lot of product measurement in the Army but not much process measurement. We assume, that because we have measured the soldier, usually having his instructor do the measurement (with all that leniency error) and not paying too much attention to standards, that we have satisfied the need to know. We know, in the Army, that programs or processes that are not examined, deteriorate. Training is no exception.

Training Program Evaluation is a "discrepancy evaluation" like system for examining the training process. It asks a series of questions about a training process, carefully stepping an evaluator through the evaluation. It examines seven aspects of training: the objectives, the design, the site, lecture events, demonstration events, practice events, and testing events. It asks some fundamental questions: Is the goal for this training specified clearly and unambiguously? Will the goal permit measurement of its attainment? Will the design for training realize the objectives? Has the design been faithfully implemented? Does the test measure the tasks that have been trained? Is the test free of contamination (hints, prompts, cues)? Are the soldiers, in fact, proficient, as defined by the objectives?

Four job aids have been prepared for this process (Kristiansen and Witmer, 1981a; Witmer 1981; Kristiansen 1981; Kristiansen and Witmer, 1981b). They have been written for use by a branch qualified Captain who is not sophisticated in educational technology.

One job aid is a guide for examining lesson plans for adherence to sound training principles. It asks twenty questions of a lesson plan (or training plan, or training outline). These questions examine, in depth, the training objectives, the training prescription (what activities have been planned for the soldier that will ensure his or her proficiency on the intermediate and the terminal objectives), and the test (does it measure what has been trained and is it designed in such a fashion that it will yield useful scores).

Another is a guide for the training observer. When we first started developing this system, we found that observing the training process is not something that everyone can easily do. Most NCOs, at least, and most educational technologists that have used the system, needed training. They could not see

"training process." What usually happens is that they get caught up in learning to do the task, or checking to see if the task is being trained correctly (training content), or they just watch without collecting any information. Since this system is data based, the observer is a data collector. The object is to capture data on what actually happened so it can be compared to what was supposed to happen. We had to develop a three day workshop for observers. The workshop consists of an introductory session where the system is explained and the training process variables described. Day two is used to actually examine or observe an ongoing training program. Day three is used to compare what the observers recorded with what the evaluator or analyst (the one running the workshop) recorded. The observers get a chance to try the system out and the evaluator gets a chance to see who is good at this and who is not. This observer's guide lists 109 questions that can be asked of the training process. Some of the items are easily seen and some are harder to see. These have been identified so that the analyst can pick the items for the observer's to use. Each item is described so that the observer knows what to look for.

Another job aid is a guide for fixing up training found deficient during a TPE. It lists a "fix" for each item in the observer's guide and has, as background for the analyst, sections describing practice and feedback. It has been described as a useful document for training developers, too, since it describes, in layman's terms, the "do's" and "don'ts" of training development and implementation.

The final job aid is a guide for the analyst, helping him or her to manage all this. It tells him or her, in detail, how to plan, organize, and carry out the TPE. It has a section on how to conduct a workshop for observers. It has a section describing how to organize to get acceptance of the findings (since the evaluator is usually in an adversary role from the very beginning). It suggests various ways to report the results.

The system has been piloted at Fort Knox against initial entry training programs and the advanced NCO course, against selected officer training modules in the basic course, against the NET program for the M1 Tank during OT-III (where three companies were trained), and against selected blocks of unit training, both individual and collective. At the M1 Tank OT-III we had to train some 100 observers and manage data from as many as 30 training sites. The system provided useful process data that would not have been available otherwise. The data was used to fix the training where it did not work very well.

A number of positive results came out of the use of TPE at the M1 OT-III. Fewer adverse comments were made by the observers during the second company's training than during the first (and the observers were independent groups). The same finding held for company three versus company two. The instructor group credited many of their program changes to feedback from the TPE. First time GO rates steadily rose from the first company to the third company. Not all of this was credited to improved training but the personnel involved in conducting the training for the OT were willing to credit TPE for much of the information on which to base training improvement.

An abbreviated version of TPE has been published in Army Trainer. It contains fifteen things to look for when evaluating training. It will be useful to units where they do not have the resources to organize an official training evaluation function.

TPE is a start toward upgrading the state of the training process in the Army. It has been business as usual too long.

References

1. Hall, Eugene R., Lam, Karen, and Bellomy, Susan G. Training Effectiveness Assessment. Vol. I. Current Military Training Evaluation Programs. TAEG Report 39, US Navy Training Analysis and Evaluation Group, Orlando, FL, December 1976.
2. Kristiansen, Donald M. A Job Aid for Modifying Ineffective or Inefficient Training Programs. Research Product 81-17, US Army Research Institute, Alexandria, VA, 1981.
3. Kristiansen, Donald M. and Witmer, Bob G. A Job Aid for the Systematic Evaluation of Lesson Plans. Research Product 81-15, US Army Research Institute, Alexandria, VA, 1981.
4. Kristiansen, Donald M. and Witmer, Bob G. Guidelines for Conducting a Training Program Evaluation (TPE). Research Product 81-18, US Army Research Institute, Alexandria, VA, 1981.
5. Provus, Malcolm. Discrepancy Evaluation. Berkeley, CA: McCutchan Publishing Co., 1971.
6. Sanders, J. R. and Cunningham, D. J. A Structure for Formative Evaluation in Product Development. Review of Educational Research, 1973, 43.
7. Schulz, Russel E., Hibbits, Nancy, Wagner, Harold, and Seidel, Robert J. On-Line Authoring Aids for Instructional Design. Technical Report 387, US Army Research Institute, Alexandria, VA, July 1979.
8. Stake, R. . The Countenance of Educational Evaluation. Teacher's College Record, 1967, 65.
9. US General Accounting Office. The Army Needs to Improve Individual Soldier Training in Its Units. FPCD-81-29, US General Accounting Office, Washington, DC, 31 March 1981.
10. Witmer, Bob G. A Job Aid for the Structured Observation of Training. Research Product 81-16, US Army Research Institute, Alexandria, VA, 1981.

ASSESSMENT OF TRAINING

The state of Army training is assessed within the framework of a general equation for trainee performance. Of the five elements of the general equation, two are identified as major contributors to enlisted training losses: leader orientation and training quality. The effects of these two elements are demonstrated through actual examples and research findings and it is suggested that their effects are largely unrecognized by the training leadership and the research community.

In this paper I aim at assessing the state of Army training in technical-skill training areas, training in those skills we would categorize as more objective on the continuum depicted in Fig. 1; generally that training, occupational training, which we administer to newly enlisted personnel. The title of Figure 1 is to highlight the essential role of measurement in evaluating training outcomes. As we approach either end of the continuum in Fig. 1, our measurements, and interchangeably, our training, become more or less objective.

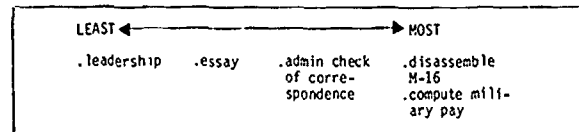


Fig. 1. Test validity continuum

The assessment in this paper is organized around the general equation in Fig. 2. Most of the data that I will show is taken from technical skill training courses for newly-enlisted Army personnel. My basic purpose is to highlight the effects of two elements of this equation which, in my opinion, are largely unrecognized by our training leadership at all levels of command and likewise are generally unrecognized by our professional research community.

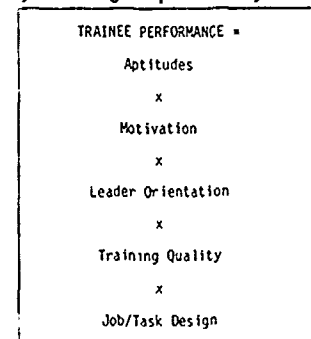


Fig. 2. General Trainee Performance Equation

The views in this paper are my personal views, distilled from 27 years association with the Army's training system as a staff training specialist, currently with an Army Service School; previously with our headquarters at Fort Monroe, Virginia and initially with another Army Service School. Obviously, my personal views do not necessarily represent the views of the Army or my current organization of association. I do have the agreement of my command to present these views. Even so, I offer these views anonymously; not because I fear reprisal, but because I believe that the criticisms I offer apply to any Army Service School (Public schools and Universities also). And, because I am certain that some will take strong issue, I don't want my school to take the "hits" alone.

With the preceding said, examine with me each element, in turn, of the general equation in Fig. 2.

Aptitudes

By aptitudes I mean the abilities measured by tests like the old Army Classification Battery (ACB), our current Armed Services Vocational Aptitude Battery (ASVAB), and similar batteries. The relationships between these such measures and various trainee performance measures has been studied exhaustively. On the average, the relationships between these aptitude measures and trainee performance measures are low, around .4, ranging up to a maximum of around .5 and to a low of .1.

We do have a few examples of considerably higher aptitudinal relationships with trainee performance. In the late fifties and early sixties several research reports by the Army Signal School at then Fort Monmouth, New Jersey reported correlations approaching .7 between the Electronic Aptitude measure of the old ACB and trainee performance. My explanation, one that you will not find in any of these research reports: strict adherence to norm-referenced testing techniques-- all training tests multiple choice, minimum standard deviations of 7, and each item required to possess statistically significant positive discrimination. A major argument within the school at the time was whether the training test cut scores should be set at minus 1, minus .75 or minus .5 standard deviation. Attrition rates in the enlisted courses were high; 20 to 40% and more. One course had a recycle rate of 120%. The generalization I draw from the Signal School research reports: you can get higher correlations between aptitude measures and trainee performance measures by norm-referencing your training tests. And, of course, you can fail a lot more students too.

An aside of interest. One of my former supervisors told me of an incident that occurred during the time period that the Signal School was rigorously applying norm-referenced evaluation techniques to their training tests and concurrently arguing for higher aptitude prerequisites. The draft was then ongoing and DA selected the training assignment of draftees based on their aptitude scores. To see what would happen, DA, unannounced, raised (considerably) the minimum required prerequisite aptitude scores for selected Signal School courses. Did the attrition rates go down? My former supervisor said that the attrition rates did not change.

Our typical trainer, at any level, gives far more credence to aptitude than is warranted. Few understand that the relationships between measures of aptitude and trainee performance are generally very low. Even fewer know that the correlations between aptitude and various measures of job performance are even more dismal, around .10, hardly worth more than a casual glance.

Our relatively recent emphasis on performance or criterion-referenced training tests has generally reduced the relationships we typically obtained in the past between aptitude measures and trainee performance. This reduction should be the goal of training and its occurrence cause for celebration. But our typical trainer generally holds opposite views.

Motivation

I reluctantly include motivation in my general trainee performance equation because our typical trainer has considerable faith that it plays a large role. "He (or she) is just not motivated," is a frequently heard remark around service schools. Questioning its reality and suggesting that the trainer can influence it borders on heresy. There are a few, nice, exceptions.

When calling on motivation to support a recommendation for failure or to explain a failure, a frequent occurrence, our trainer is usually referring to internal or intrinsic motivation. References to external or extrinsic motivation are rare. The distinction between the two forms, and the trainer's role in the latter, is largely unrecognized in the context of training.

Commanders and directors of training have, with some frequency, ordered the development and application of motivation tests and, we have used them to counsel students and to decide their disposition. We did not ask, we did not know to ask, of the validity of the test.

In fact, we can't measure either form of motivation with any degree of acceptable accuracy for decision making about individuals. But, we can demonstrate with impeccable evidence the existence of extrinsic motivation. In a context now disassociated from training, both trainers and non-trainers across the Army believe in extrinsic motivation. (More on this in the next section.)

Leader Orientation

This is the least emotionally loaded label I could conceive of to denote what I believe to be a most significant factor in the trainee performance equation. Here are several examples, representative of many more, which collectively suggest that leader orientation has great influence on the trainee performance equation.

The attrition rates in Table 1 are taken from US-CONARC publications (US-CONARC was the predecessor of our current Training and Doctrine Command). These were annual publications entitled "US Army Service School Training Reports." All three of these courses were of the same length and all three aimed at training newly enlisted personnel to perform field maintenance of the identical field radio. Trainee aptitude prerequisites were the same for the three courses. Except where noted, all three were large volume courses; 500 to The Armor School had an article published in "Army" explaining how they did it; basically "hands-on" practice and lots of it,

TABLE 1
Loss Rates (in Percent) in the Radio Communication Courses
Conducted by Infantry, Armor, and Artillery Schools. *

FY	INF	ARM	ARTY
65	18.1	.9	11.5
64	11.7	6.0	29.0
63	12.2	22.5	36.0
62	5.8	19.0	28.0
61	5.2	17.0	16.5
60	8.7	11.4	-- ***
59	15.5	0.0**	12.6
58	19.2	14.1	14.8
57	17.8	8.2	20.7
56	24.7	8.9	25.0

* Data from Fiscal Year USCONARC School Course Attrition Reports. Output each FY was volume training, 5 to 24 hundred except where noted.
** Output was 53; no losses.
*** No training conducted.

coupled with a leader orientation of "You send them, we'll train them." The other two schools continued with multiple-choice testing and their long standing concern with "poor quality of students." Our training leadership at the headquarters believed that an "M1 Pencil" (A WWII expression to indicate falsifying test scores) explained the small attrition rate at the Armor School: They were passing people they should have failed. But field supervisors were unanimous in their praise of the graduates. Leader orientation can explain much of these differences.

Table 2 illustrates a more recent example of leader orientation. These disproportionately high female and male loss rates occurred at a time when female aptitude prerequisites were much higher than male aptitude prerequisites. The differing loss rates cannot be explained by conceding sex differences in math (if you have such a persuasion . . . research has not resolved the issue . . . it is a current hot debate topic in the literature). Some of the differences are quite large and statistically significant by Chi square tests (see table 2). I prefer the explanation of sex bias against the

females in courses A and F. If aptitudes mean anything, the ladies should have outperformed the men. At the time, the ladies had to meet much higher course prerequisite aptitude standards than did the men. The commandant of the school elected sex bias as the explanation and the differences in the failure rates ceased following the Commandant's discussion with the course directors involved. It was rather difficult to argue that particular courses were being deliberately assigned inferior female (or male) trainees in comparison with other courses.

Table 3 provides an example of significant loss differences among components (Active Army, Reserve, and National Guard) in one course; clearly different from component loss rates in other enlisted courses. The differences were recognized when the Commandant asked why loss rates were up in that particular course and the instructors replied that the National Guard students were the cause. The reported aptitude scores of the National Guard students were alleged by the instructors to be considerably inflated: at the time National Guard personnel administered the ASVAB to their recruits.

TABLE 2
Male and Female Loss Rates in Eight Enlisted Courses *, FY 77 and FY 78

Course	N		% Loss		X2**
	Male	Female	Male	Female	
A	78	45	15.4	35.6	6.604***
B	1,439	426	19.5	7.3	35.410****
C	459	478	14.6	12.6	.836
D	282	166	11.7	9.6	.457
E	213	325	20.2	27.1	3.315
F	170	88	27.6	39.8	3.932***
G	465	624	8.2	8.3	.009

* Minimum mental standards for female accessions, as measured by the Armed Forces Qualification Test (a general screening test) was at the time much higher, 59th percentile, than for males (between 15th and 31st percentiles).

** Chi Square tests for differences between the numbers of male and female students failing.

*** $p < .05$

**** $p < .005$

TABLE 3
Comparison of Loss Rates Among Components in the PASC and Other Enlisted Courses*

Component	COURSE "A" **		% Loss Other Courses ***
	N	Loss	
Regular Army	354	4.5	15.3
Army Reserve	119	12.6	17.1
National Guard	402	20.6	16.7

* Data from first 10 months of FY 79.

** A test of differences among component loss rates yielded a chi square of 43.23, $p < .005$

*** About 6,000 students in about the same component proportions as Course "A"

As the reported aptitude scores were alleged to be false, testing them for differences was meaningless. Aptitudinal differences among the three components evaluated via a different aptitude test locally administered. There were no statistically significant differences among the aptitudes of the three components.

The instructors then suggested that National Guard students were not responding to a large portion of the course dealing with what was called SIDPERS: the Army's automated personnel management system. Personnel management in National Guard units was different from SIDPERS: thus, it was alleged that National Guard students were indifferent to, and therefore failing, the SIDPERS portion of the training. A large sample of test scores from the SIDPERS training were tested for component differences. No statistically significant differences were found.

Several other possible explanations were checked. The last one checked identified what came to be, reluctantly, the accepted explanation: a general instructor bias against the National Guard student; the "Weekend Warrior," and other choice expletives. Brought to the surface and gently discussed with the instructors, the differing loss rates among the components ceased.

Table 4 illustrates racial loss differences in one course quite divergent from the long accepted directions. I offer an explanation that I must carefully qualify. It is speculation on my part: I readily admit that I cannot prove that my explanation is correct. This exception to the expected racial loss rates coincided with the assignment of a black leader as course director. In no way do I suggest that his leadership orientation was consciously biased. Rather, his strong, laudable, orientation with black students was: "You are not going to fail. We must demonstrate that we can." From observing this leader interacting with black students, one would agree with the descriptive labels of "dynamic and charismatic," in his dealings with them. High expectations of success! But unknowingly aimed at only one category of student. With the normal rotation/replacement of this course leader the figures changed to their traditional relative proportions.

TABLE 4
Majority and minority loss rates in one course and seven other courses.

Time*	Student Category	% Loss (N)	
		Course A	Other Courses
802 & 803	Majority	3.5 (201)	10.1 (975)
	Minority	6.7 (239)	13.6 (982)
804, 811 812, 813	Majority	3.6 (614)	11.9 (3,306)
	Minority	2.4 (660)	18.7 (2,569)

* The numbers "802, 803", etc., indicate, respectively 2nd Quarter FY 80, 3rd Quarter FY 80 etc.

Thus far we have examined several examples of what I believe can be fairly called leader orientation effects within particular courses.

The influence of more senior levels of leader orientation are, I believe, also observable. Figure 3 displays attritional losses, initially by FY, then by quarter, in enlisted courses over several years within my school.

The nine enlisted courses are all lumped together in this figure and the total numbers are large; around 8,000 each year and around 2,000 each quarter. Of interest here is what explains the "ups and downs"? The explanations usually offered revolve around "input quality." Even "seasonal variations in input quality" has been offered as explanation. I think that senior leader orientation explains a great deal of the variation in attrition evident in Figure 3. Here are the facts that lead me to this personal conclusion.

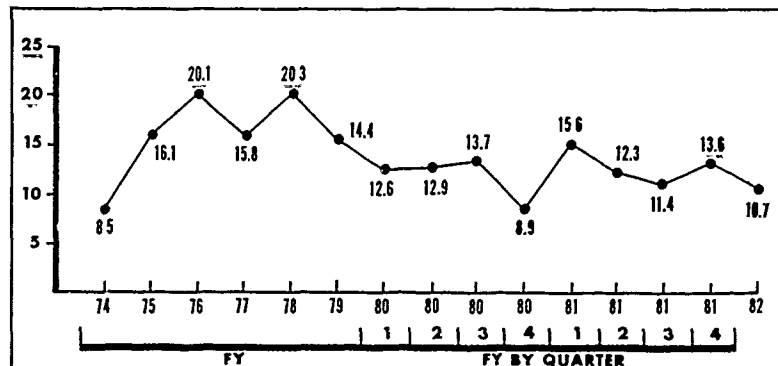


Figure 3. Enlisted course losses in percent

From FY 67 through FY 75 the enlisted course loss rates taken together ranged from 8 to 12 percent. The Army Service Schools underwent a major reorganization in FY 75. One significant result of this reorganization locally, as related to attritional losses, was the elimination of a five man staff element whose full time, and sole, mission was monitoring training quality through training test results. Beginning in FY 76, there was no command emphasis on training quality (and I recognize that the term "training quality" was then as now, quite controversial in meaning and practice). Through the 77-78 FY time period the enlisted course attrition rates reported during each Command Performance Summary were incorrect. (The data in Figure 3 are the correct loss rates.) The reported rates each quarter for each of the nine enlisted courses were, on the average, about half of the true rates of attrition. I will never know whether this under-reporting was accidental or deliberate. I know how it occurred: TRADOC changed the 20 year old formula for computing attrition rates and the formula was either deliberately or unintentionally interpreted to produce the under-reports. For about a two year period our senior leadership reviewed attrition reports which under-represented the true rates of attrition by nearly half. There was no command concern with attrition during this time period. You must judge for yourself whether or not the high attrition points can be explained, at least in part, by command's believe that attrition was lower than it actually was. I believe this explains much of the differences.

Late in FY 78 I assumed a staff position within the resident training side of my school, discovered the under-reporting, and advised the newly assigned Commandant. This new Commandant had already expressed great concern with the high attrition rates. This concern intensified to the training staff on the discovery of the under-reporting. I believe this explains the significant drop in attrition rates observable beginning in FY 79.

A new Commandant, assigned in the 3d Quarter of FY 80 ("80₃" in Figure 3), continued the previous Commandant's concern with enlisted course rates of attrition. When the 4th Quarter (80₄) enlisted course attrition rates were reported, this Commandant expressed concern with the quality of the graduates. Were we lowering standards? The 8.9% loss rate was the lowest ever. I believe that the dramatic jump in the enlisted course attrition rates, up to 15.6 percent for Quarter 81₁, was directly a result of this expressed command concern with quality and standards. On seeing the 1st Quarter FY 81 rate, our Commandant suggested that we may have gone too far in emphasizing standards and quality. And down the figures went!

The lumping together of all of the enlisted courses in Figure 3 masks still another observation. These individual course attrition rates fluctuate erratically over time just as was observable previously in Table 1: effects of leader orientation.

A former senior training leader within my school commented on a draft of this paper:

I have a thought or two on your figure 3 and the accompanying discussion. As you know . . . I could give you any attrition rate you wanted and rationalize my way clear. Reason: Because I would (today anyway) control all the testing and all the scoring, and there is an ample supply of "M-1 pencils" to meet any "acceptable attrition rate."

Several years ago a general officer letter addressed one of our NCO courses which was of great concern to his command. His letter pointed out that our NCO course had an attrition rate that was one-half to one-third lower than similar other-service courses and suggested that our standards were too low. Our NCO course attrition rate promptly increased, in line with other-service course attrition rates. This, despite the fact that we, and his command, had independently of each other studied the relationships between measures of school and job performance and had found no relationship. (Neither of the two studies were published.)

In passing, the many formal studies of the relationships between performance in training (formal school settings in universities, businesses, industries, and military schools) and various measures of job performance are entirely consistent in their results. The two are not related. This is quite a shock to most military training leaders. One of my former commanders almost threw me out of his office on hearing this assertion. To his credit he later accepted the generalization and based thereon vigorously pursued elimination of "class rankings" and designation of "honor graduates." We did stop the practice . . . until he left.

Recall the earlier discussion of motivation. We are referring to extrinsic motivation when we talk about leaders influencing the performance of their units. I don't like to work with figures (intrinsic motivation) but my leader has convinced me that we must get these computations done (extrinsic motivation). Our typical trainer believes in extrinsic motivation in the context of unit performance but has little respect for the notion within the context of training.

That instructors (leaders) influence extrinsic motivation within training settings has been well demonstrated by research. Rosenthal and Jacobson (1968) administered a test of "potential" to 650 grammar school children which identified about twenty percent as likely to demonstrate academic growth above expectation the next academic year. The names of these potential academic "spurters" were given to the teachers. The "potential" test did the job! Those children so identified did achieve above expectation. The catch to this research: the test scores were ignored and the potential spurters were selected by a table of random numbers. How an instructor perceives a student does have impact! According to one writer, this research, now somewhat of a classic, has been duplicated over a dozen times in a variety of instructional settings.

Our military training leadership, at all levels, is in general quite insensitive to the leader orientation portion of the trainee performance equation. While we have great respect for leadership in units, and we invest significant resources in leadership training, we are not really conscious of its operational effects in our training. When there is concern for too high attrition, almost without exception the response explanations involve aptitudes.

Aside from particular biases observable (sex, race, component, etc.), there are two basic leader views of the role of training which predominate: a "selective" and an "instructional" role. The former view (select out the unfit) is more frequently observable. You hear this former view in infinite and subtle variations. Fig. 4 illustrates a few of the frequently heard comments around service schools which reflect a selective orientation. The other basic orientation, "you are trainable," is observable but, in my view, too rarely. This "failure" orientation has its origins in our educational culture. Bell curve grading is our cultural norm. We have had this indoctrination, from kindergarten on: there are mostly average and a few good, and bad, students; failure is alright, even expected, things are not right unless there are failures (the course is too easy). Mastery or task competence is expected only of a few. Achieving seventy percent is "passing." Average . . . mediocrity . . . is our norm.

1. "The field is yelling about the quality we are sending them. We gotta tighten standards."
2. "What are you trying to do, spoon feed them?"
3. "You ought to look at your standards, your attrition is so low."
4. "That attrition is about right."
5. "You have just completed a most rigorous and demanding course of instruction. Its difficulty can be measured in some degree by the number who began the course with you but are not graduating with you here today." (Extract of script for the graduation ceremony of an enlisted course.)
6. "Most of our students come back at night for long hours. It's a tough course!" (With pride.)
7. "I wouldn't have that student working for me . . ." "We gotta . . ."
8. "That is a tough test."

Figure 4. Quotes illustrative of a selective orientation versus an instructional orientation.

Training Quality

Training quality is that element of the trainee performance equation which is, in my opinion, a most controversial issue and one that is greatly bound by our educational experiences/cultural expectations. I also believe it to be the element with the most potential for gains.

Our culture has towering faith in education and training. In dollar value we devote nearly eight percent of our gross national product to public education and about the same proportion of the defense dollar to formal training efforts. At any one time, about 15% of our total Army strength is either in our schools, conducting school, or supporting our schools. We do have faith that training makes a difference!

Notwithstanding our enormous belief in the value of training we have no consensus on what is quality training. From this point on, the water I am wading in gets deeper and more troubled.

I am going to take one example which I believe, had you been with me during a recent ten year period, you would agree is typical and representative of literally hundreds of similarly observable incidents within my school.

The "task" is taken from a 10-week course with very heavy demands on the higher-order verbal skills of listening and writing. In this task, our trainee must, via his or her long hand notes, produce a typewritten summary of dialogue typically with around eight participants and ranging in duration to an hour or more. Tape recordings are not allowed. A typical product produced by a student would run 20 pages, about half forms, with the rest a summary of the dialogue; very detailed, bordering on a verbatim transcript.

When the training was designed, there was a great deal of emphasis on what has subsequently come to be called "front-end-analysis" with the particular, even singular, objective of designing the training test, before the training was designed. Even today this is considered to be a somewhat radical approach.

The test took the form of a video tape, scripted and produced at considerable cost (and still in use to this day). Scoring was by a check list of over 300 elements; a date, a time, a place, what was said and who said it: each element scored by product inspection as right or wrong. What the student viewed as a single test was actually managed as three tests.

Table 5 displays, for the forms portion of the test (the least difficult of the three parts), the frequency distributions resulting from the 1st and 16th class. Note the significant difference between the two distributions; the first normal, the second highly skewed. The distributional shapes in between the first and 16th classes each became progressively more skewed. A graph of the means of the test scores over the 16 classes would approximate the classical learning curve. Note that most of the scores in the second distribution are above the median score of the first distribution.

TABLE 5
Distribution of Test Scores Resulting from the First and Sixteenth Classes.

Raw Score	1st	16th
56		////
55		////
54		////
53	///	////
52		///
51	///	///
50	///	///
49	///	///
48	///	///
47	///	///
46	///	///
45	///	///
44	///	///
43	///	///
42	///	///
41	///	///
40	///	///
39	///	///

* Cut score applied for the first class.
** Cut score applied for the 16th and all subsequent classes.

After each test, item analysis was performed in, then, quite a new fashion. An entirely new statistic became almost the sole basis for item analysis and training quality control. We labeled this statistic "percent item miss rate." We insisted on, successfully in this case, and in hundreds of other cases, one single rule for its interpretation. For item miss rates of around 20% or more: suspect, find, and fix an error in the training design/delivery. The principle was: training design must be good enough for most students. Its operational definition was: item miss rates of around 20% and higher indicated that the training was not good enough. Operating under this principle the instructional cadre shaped, molded, and redesigned this unit of training. Simple design errors were identified and corrected: An incorrect illustration of a signature block in the training materials, when fixed, eliminated a 50% miss rate. A 90% miss rate was attributed to a portion of the video tape (the media of the test) being simply inaudible -- you couldn't hear what was said. A 40% miss rate was traced to scoring disagreements. Scoring unreliability was rather large initially as one might suspect.

But as succeeding classes passed, the by-far larger problem became increasingly clear. There was simply not enough practice. Over the first year the practice materials and the classroom training time devoted to them actually doubled. It took over a year to develop the needed additional practice materials and two additional alternate test versions. The three alternate test versions were then used randomly and interchangeably for practice, for grade, and for makeup/retesting. The original estimate of 6 days of practice proved to be only half right.

From this one example and, accept as true, hundreds of similar examples in other courses, I am now in a position to make several observations.

The largest single factor in the trainee performance equation is, on the average, "training quality" or more specifically, the lack of training quality.

We should redefine our classical model of a test score as illustrated in Fig 5. The error portion of the true score is typically presented in our measurement literature as stemming solely from random errors and the state of the learner; he didn't feel good when he took the test and his observed score is less or more than his true score, etc. The inference in the classical model is that there is nothing inherently wrong in the test, test administration, test scoring, training design, leadership, and training administration. But, there usually is!

$$S_t = S_o + S_e + S_d$$

Where:

- S_t is true score
- S_o is observed score
- S_e is observed score (measurement error)
- S_d is error score (design, delivery, leader orientation error)

Fig. 5. Revised Model of a Test Score.

Our educational and training cultures, both public and military, simply do not comprehend that test item difficulty is, in large part (I would say "for the most part"), a function of training design, leadership, and administration.

Criterion-referenced notions have had some impact but in the main our typical trainer and educator remains largely unaffected by them. They continue on serenely: "70% is passing;" "that is a difficult test." Our professional literature, read, if at all, mostly by other professionals, is viewed by trainers and teachers as a foreign language, full of esoteric terms. Questions of fidelity, transfer of training, reliability, skill decay, etc., keep the waters of our professional literature in constant agitation while the testing waters of our schools remain largely placid. These technical terms are mostly "non-terms" to the vast majority of our trainers. Try dropping the terms "psychological fidelity" and "stimulus generalization," both representing very powerful and useful concepts, into a conversation with the typical trainer and see how long the conversation lasts. Suggest that "The test is OK, let's beef up the training," and watch the eyes roll. Advise that the cut score on a test should be set initially in consideration of training quality (usually very poor) and subsequently adjusted (up) as the training quality is improved, and observe the knowing side-glances.

Job/Task Design

I mean, by the term "job/task design," the configuration of the occupational specialty; how we prescribed the performance of each of the tasks, and how we designed the job givens (manuals, tools, forms, etc.). We have some job tasks even the most optimum training design has great difficulty in dealing with.

Two examples to illustrate the issue here. Our original Prefix 5 training, nuclear weapons employment training, was heavily loaded with some pretty esoteric math -- we brought in physicists to train our people in nuclear weapons employment; compute fall-out ratios, decay rates, etc. Much of the early learning difficulty was a function of how we had defined task performance. Simple nomographs later eliminated all the learning problems experienced by students with minimal mathematical training (and there were quite a few). The nomographs also reduced the learning time and increased task performance reliability. While illustrating a magnificent improvement in the training, the story also illustrates that we can do a sorry job of designing a task.

The second example. Back in the days of the 125-item multiple choice test for enlisted evaluation (did we really get away from that?) there was a carbon set form (one of many) that our personnel clerks had to complete which identified personnel eligible for enlisted evaluation (an interesting word choice . . . you weren't just "eligible", you had to). We were going through a task analysis of the task requirements associated with enlisted evaluation. In the process we stumbled into over 30 design errors related to one carbon set form. A few examples. The form (with tractor holes along each side for ADP processing) was too wide for the standard Army issue typewriter. Tear off the perforated tractor guides to get the form into a typewriter and you had problems keeping the carbon copies aligned. The sequence of information required on the form was not the same as the sequence in the source document -- more chances for error, more time required, than necessary.

The two preceding examples are agreed upon history. Depending on our perspective, we call them examples of failing to apply "forms management," or "human engineering." From still another perspective, we didn't think of applying "job aid" concepts. All of the labels deal basically with the same concern: designing a task/tools/manuals, etc., in such a fashion that the learning requirement is minimal. Going back to the mid 20's, numerous research studies have addressed the job/task design issues, many of them under the rubric of "job aids" or "job performance aids." Using a variety of formants, study after study has reported whopping gains in reduced learning time, greater accuracy, etc. The message I get from this really large body of research is that we can achieve large savings through the application of these techniques. The other side of this message: how much weight does the element of job/task design have in the trainee performance equation? It has significant weight, more here and less there. The research clearly supports the existence of this element of the equation. Even the most optimum training design, leadership, and administration cannot overcome a poorly designed job/task.

A possible current example, one with which most of you are likely familiar and have probably felt victimized by; the Joint Travel Regulations. Only the chosen few really understand that monster and I'm even doubtful of them. Imagine trying to train Clerks, newly-enlisted personnel, to perform in that arena.

Again, our training leadership is little sensitized to this potential source of trainee performance variance. The job-aid advocates within the research community are quite sensitized but they are a small group and rarely heard by our training leadership.

The job/task design element of the equation is, in the immediate time frame, a given for the trainer. The trainer must try, through design, to compensate for a poorly designed system/task. I don't have a strong opinion on how will we do this, but, I suspect that we don't do it very well and that such is often the root source of field performance problems and related field commander gripes: to which, our schools are usually quite responsive but not necessarily in the correct direction.

Summary

Of the five elements of the general trainee performance equation, our training and research leadership should focus on the element of training quality. Acquiring and applying a new definition of training quality ("the training must be good enough for most students") we would then target on that element offering the most gain. By such emphasis, we cease our generally fruitless efforts with aptitudes and intrinsic motivation. We stop our explanations ("naming" is not explaining) that a particular test is "hard," that a particular test question is a real "ego buster," that particular classes are bad, that a 20 (or 40, or 60) percent attrition is about right. We stop judging training by how it looks; a focus on the processes of training. Instead we focus on the products of training. This product focus is not a myopic

"we-love-everybody", "everybody-is-good" approach. It is a hard nosed, patient, long-term, empirically based approach; a systematic, methodical trial-test-revise approach to the design and management of training: find, and fix the (inevitable) training design and delivery errors!

By focusing on training quality (let's figure out how to . . .) we subtly, but by the most direct possible means, attack leader orientation; the equally important element of the trainee performance equation. By such means we attack, and change, our basic cultural attitudes of failure and poor quality of students. We stop accepting 20, 40, and 60 percent attrition rates.

Where should our training research effort go (big bucks here)? Not on the first two elements. That has been done and done. We know the answers here. On the last element; job/task design? Yes, continue effort here but be a little humble. Poorly designed jobs and tasks will continue to be implemented under operational contingencies. Yes, yell about them loudly! But recognize that our typical military trainer is around too short a time to place his marbles on such a long-term pay off. And, recognize that our typical civilian trainer in a military service school has vested interest in that task. He will lose that 40-hour block of instruction which is why he really exists (only he really understands it) when you eliminate the learning problem through re-designing the task.

We lose, on the average, year after year, over one out of four of our new accessions, new enlisted personnel, through Basic Training and Advanced Individual Training. The vast majority of these losses are not due to mental, physical, or disciplinary reasons; they are academic losses. They simply can't pass our tests. Surely, leader orientation and training quality account for a significant part of that loss. Reason enough to focus heat on the training quality and leader orientation elements of the trainee performance equation.

"MARK HOPKINS"

RELATION OF COOPERATIVE RESPONSE TO INTERPERSONAL NEEDS

JAREAN RAY
AND
PAUL BROWN*

UNITED STATES AIR FORCE ACADEMY, COLORADO

ABSTRACT

This paper investigates the amount of correlation between cooperative behavior (as measured by the Prisoner's Dilemma game) and inclusion behavior measured by the FIRO-B test, as demonstrated by cadets at the United States Air Force Academy.

PROBLEM

The principles of game theory were utilized to model a situation of conflict. In this situation, the subjects chose a response that demonstrated an attempt to cooperate or a response that indicated other motives overriding cooperation. The FIRO-B test measured behavior based on interpersonal needs for each subject. The hypotheses are as follow:

H₁: The higher a person's score for Inclusion (Wanted) the more likely he/she¹ will select a cooperative response.

H₂: The higher a person's score for Inclusion (Expressed) the more likely² he/she will select a cooperative response.

H₃: The higher a person's combined score for Inclusion (Expressed + Wanted)³ the more likely he/she will select a cooperative response.

METHOD

SAMPLE

The subjects were 26 volunteers selected at random from cadets who were studying at the Academy library.

PROCEDURE

Each subject received a brief, general explanation of the experiment; to keep the explanation from influencing the responses, the subject was told only that the purpose of the experiment was to determine the amount of correlation between responses on two different surveys, and that he/she would receive an explanation of the results for his responses after the experiment. Because the questions on the FIRO-B test might have caused the subject to be in a more cooperative frame of mind, the subject was required to complete the Prisoner's Dilemma prior to the presentation of the FIRO-B test.

*Requests for reprints should be addressed to the second author.

INSTRUMENTS

The Prisoner's Dilemma

The Prisoner's Dilemma game produced the conflict in this experiment. It modeled the situation in which two suspects were faced with the choice of confessing or not confessing to a serious crime. The subjects were instructed that they did, in fact, commit this major crime. They are separated, with no communication allowed. The authorities already have enough evidence to convict each suspect of a lesser crime but are willing to deal with the suspects. In order to get a conviction of the major crime, the authorities offer "freedom" to the suspect who confesses if the other does not confess, while the other would receive the maximum prison sentence. If both confess, both will be convicted of the major crime and receive a sentence only slightly less than the maximum sentence. If neither suspect confesses, they will both be convicted of the lesser crime and receive a light sentence.

According to the game theoretician's definition, a rational decision would entail each suspect acting in self-interest to receive the least time in prison (the highest-valued payoff). Hence, it seems that the best strategy would be to confess. If suspect A chooses not to confess, then he may get the light sentence, if suspect B also chooses not to confess. However, if suspect B chooses to confess, suspect A will receive the maximum sentence. On the other hand, if suspect A chooses to confess, he will either receive the reduced sentence or be released, depending on subject B's decision. It seems, then, that the safest course of action for suspect A is to confess, so that he will avoid the maximum sentence. This argument is especially appealing when considering the most logical view that suspect B will tend to use the same reasoning, also choosing to confess, so that neither would be released. The dilemma is evident when one considers that the resulting prison sentence is more than each would have received had neither confessed (Rapport, 1960).

The FIRO-B Test

The acronym FIRO stands for Fundamental Interpersonal Relations Orientation. The B after the dash indicates the version that measures expressed and wanted behavior. Developed by Will Schutz, Ph.D., the FIRO theory proposes that interpersonal behavior is determined by the characteristic ways every person orients himself toward others (Shaw and Costanzo, 1970). The theory states that interpersonal interaction is a result of three interpersonal needs--inclusion, control, and affection--that develop during childhood (Shaw and Castanzo, 1970). Inclusion refers to integration into a group, control involves the power and responsibility in decisions between people, and affection refers to the emotions between people (Shaw and Costanzo, 1970).

For each interpersonal need, the FIRO-B test measures the behavior a person expresses and the behavior a person wants from others, assigning a score ranging from zero to nine with a higher score indicating a higher need. Since both expressed and wanted behavior are interrelated and may influence a person's decision, the combined score (E + W) for each need may prove to be a more accurate indicator of a willingness to cooperate. The FIRO-B has been

used in numerous research studies and has been found to be a very valid instrument in identifying expressed and wanted behavior for the three interpersonal need areas (inclusion, control and affection).

ANALYSIS AND RESULTS

The data was analyzed to determine the relationship between a cooperative response and all three need areas. A regression was run on three sets of data, each with different values for the interpersonal needs: data for (1) expressed behavior, (2) wanted behavior, and (3) the sum of expressed and wanted behavior. In all three cases, the linear model used the response to the Prisoner's Dilemma (a value of zero for "confess" and a value of one for "not confess") as the observation resulting from the scores for interpersonal needs.

Correlation

A summary of the correlations between the interpersonal needs and the response "not confess" is given in Table 1.

TABLE 1 CORRELATION WITH RESPONSE "NOT CONFESS"

BEHAVIOR	INCLUSION	CONTROL	AFFECTION
expressed	0.13496	0.17030	-0.05774
expressed + wanted	0.25560	-0.03360	0.08352
wanted	0.29019	-0.28778	0.21066

For expressed behavior, the correlation indicates that there is almost no relationship between the three interpersonal need areas and the response to the Prisoner's Dilemma.

For the sum of expressed and wanted behavior, there is nearly zero correlation between the response to the Dilemma and the need areas of control and affection. However, there is a small amount of correlation between the response and the need area of inclusion. Of the three areas, inclusion is the one that will most likely predict or explain the response of "confess" or "not confess" (Whalen, 1981). The correlation is positive, indicating the greater a person's expressed and wanted inclusion are, the more likely he will be to choose the response of not confessing. Although expressed plus wanted inclusion can be used to predict or explain a response to the Prisoner's Dilemma to a certain extent, other factors must also be considered (indicated by the fact that the correlation is not very high).

For wanted behavior, inclusion is again the interpersonal need behavior that has the greatest relationship with the cooperative response. Again, the positive correlation indicates that the greater a person's desire for inclusion the greater his tendency to choose not to confess. The correlation of the cooperative response with control is nearly the same as with inclusion, only negative. Therefore, the lower a person's desire for control, the greater

his tendency to choose the "not to confess" option. The correlation between a response of "confess" and desire for control will be a positive value with the same magnitude. This indicates that the greater a person's desire for control, the more likely he will be to choose a response of "confess". Affection also has a slight positive correlation with the response of "not confess". Its magnitude, however, is less than that of the other two need areas. This indicates that a desire for affection will have less of an effect than a desire for inclusion or control.

SIGNIFICANCE

The partial F statistics for each variable (need area) will be used to evaluate their significance in predicting or explaining the cooperative response. If the partial F statistic of a variable is greater than a control statistic, F_{in} , the minimum value for a variable to be added to the model, then that variable is significant in predicting the response. The value for F_{in} comes from a table of values for the F distribution, using $F_{1, 24}$, $F_{2, 23}$, and $F_{3, 22}$ respectively, at a level of significance of $\alpha = .05$ (Hines and Montgomery, 1980). For expressed behavior, the results are shown in Table 2.

TABLE 2 SIGNIFICANCE OF AREAS FOR EXPRESSED BEHAVIOR

Variables in Model	Partial F	F_{in}
Control	0.717	4.26
Control Inclusion	0.807 0.546	3.42
Control Inclusion Affection	0.459 0.565 0.053	3.05

None of the variables for expressed behavior have a partial F statistic greater than F_{in} for any of the three models. This indicates that none of the variables are significant. Hence, the interpersonal needs for expressed behavior cannot be used to predict a response to the Prisoner's Dilemma in any way. This conclusion is consistent with that indicated by the correlations.

For wanted behavior, the results are given in Table 3.

TABLE 3 SIGNIFICANCE OF AREAS FOR WANTED BEHAVIOR

Variables in Model	Partial F	F _{in}
Inclusion	2.207	4.26
Inclusion Control	5.331 5.288	3.42
Inclusion Control Affection	2.846 5.077 0.055	3.05

Although no variable has a partial F statistic large enough for it to be the only variable in the model, the partial F statistics for inclusion and control show that they are significant when the model contains only these two variables. The variable affection has an extremely small partial F statistic, indicating that it is not significant. Therefore, affection cannot be used in the prediction or explanation of the response to the Prisoner's Dilemma. These findings indicate that wanted inclusion and control can be used to predict whether a person's response will be to confess or not to confess. The correlations imply that the larger the desire for inclusion and the smaller the desire for control, the greater the tendency will be to respond "not confess". Although significant, the low partial F statistics and low correlations suggest that other factors are involved.

For the sum of expressed and wanted behavior, the results are given in Table 3.

TABLE 4 SIGNIFICANCE OF AREAS FOR EXPRESSED + WANTED BEHAVIOR

Variables in Model	Partial F	F _{in}
Inclusion	1.678	4.26
Inclusion Affection	1.638 0.190	3.42
Inclusion Affection Control	1.976 0.451 0.384	3.05

Again, none of the variables have partial F statistics greater than F_{in}. Therefore, for the sum of expressed and wanted behavior, none of the interpersonal need areas contribute significantly to the prediction or description of a cooperative response to the Prisoner's Dilemma, despite the fact that there was a small amount of correlation between the cooperative response and inclusion.

DISCUSSION

These results indicate that a person with a greater desire to be included by others and a smaller desire to exercise control over others will be more likely to choose to cooperate by not confessing. It seems reasonable to conclude that the desire to be included would cause a person to cooperate, this subject will risk being double-crossed, especially if he/she had only a slight desire to exercise control and would not be as concerned with trying to double-cross his/her partner.

The results further indicate that wanted behavior--not expressed or even a combination (E+W) is most correlated with the cooperation response. A person is much more likely to make a decision based on his desires for behavior than on his expression of behavior. A person always has the free will to choose his course of action, and he most often does so on the basis of what he wants (whether to gain something or to avoid consequences). This does not imply that his choice is based purely on self-interest, since he may choose a course of action because he wants to benefit another person.

One assumption of the experiment is that the response of not confessing represents an attempt to cooperate. This does not imply that the response of confessing represents competition. Examining the two responses, one can delineate many motives for confessing. First, a prisoner may choose to confess in order to punish his partner if he expects that his partner will not confess. In this case, his partner will receive the maximum possible prison sentence. Second, a prisoner may choose to confess, hoping that his partner will not, in order to gain his freedom. The prisoner is more concerned about the chance of being released than he is about the fact that his partner will then receive the maximum sentence. Third, a prisoner may choose to confess in order to play it safe. If he confesses he will not receive the maximum sentence, no matter what his partner decides. Fourth, the prisoner may choose to confess out of ethical considerations. After having a change of heart, knowing he has done something wrong, he may choose the morally right course of action of admitting his wrong and receiving his punishment. Clearly, there are many different reasons for choosing the strategy of confessing, and not all reasons indicate competition. However, if the prisoner decides not to confess, his overriding concern is something other than punishment, freedom, safety, or ethics. The assumption is that this concern is an attempt to cooperate with his partner for the lightest sentence for both of them (Shaw and Castanzo, 1970). Considering the numerous reasons for choosing the response of confessing, it is surprising the large number of subjects who responded, "not confess".

Thus, data indicates that many factors are involved in the "not confess" response, as well. The large number of cooperative responses by cadets is probably due to a great extent to the environment of the United States Air Force Academy. First of all, the atmosphere of living so close together in the cadet area; the great degree to which ideals, goals, and beliefs are shared among cadets; and the common experiences of the Academy life all promote a knowledge of fellow cadets, and a feeling of similarity. These factors make it much easier to imagine that an accomplice in the Prisoner's Dilemma would tend to act in much the same way as the subject himself. Another

factor that would add to this reasoning is the great amount of control exercised over cadets, since this control allows few deviations from norms. The fact that there are not many deviations from norms would also promote a conclusion that an accomplice would make the same decision (Linden, 1981).

Another factor of the cadet environment is living under the Honor Code. Such an experience tends to promote trust of fellow cadets and people in general. A greater amount of trust would promote a decision to cooperate in the face of risk--the risk of being double-crossed would seem small if the subject perceived mutual trust.

Furthermore, the competitive environment of the Academy promotes the desire to achieve the best possible outcome relative to a "value of self." This desire would correspond to a desire to receive the lightest sentence as an overriding motivation toward a decision. The practice of "gaming" at the Academy (getting by with as much as possible without getting caught) would also make the decision of not confessing appealing to cadets. The subjects will risk a double-cross in order to receive the lightest sentence. This high risk situation is very similar to the way many cadets practice "gaming the rules."

LIMITATIONS

There are several limitations which may have produced biasing effect in this experiment. For example, the sample size and the administration of the experiment required that the subjects be chosen on a voluntary basis. Thus, the sample will have some amount of bias due to the method used for choosing the subjects, and it may not be as representative as a larger sample chosen strictly at random.

Finally, it was assumed that other experimental controls were adequate. In other words, no bias existed in the way it was administered, the wording of the survey, or in the time and location of the subjects as they participated in the experiment. When examined before the experiment was conducted, these factors did not appear to be significant. However, the fact that the administrator was aware of the different responses as they were being collected may have influenced the experiment. Also, the level of significance, $\alpha = .05$, was not set before the experiment began. This does not change the fact that the results are significant at this level.

CONCLUSION

The purpose of the experiment was to find what correlation existed between a cooperative response to the Prisoner's Dilemma and behaviors in the interpersonal need areas described by the FIRO theory, as demonstrated by Air Force Academy cadets. The hypothesis was that a person with a greater need for inclusion behavior would be more likely to choose a cooperative response (the additional correlations were determined ex-post facto). The experiment determined the following points:

- (1) There is a small amount of positive correlation between a response of "not confess" and wanted inclusion.
- (2) There is a similar amount of negative correlation between a response of "not confess" and wanted control.

(3) A desire for inclusion and control are significant factors in predicting or explaining a response to the Prisoner's Dilemma.

(4) Other factors must be controlled when predicting or explaining a response to the Prisoner's Dilemma.

On the basis of these results, only the first hypothesis was supported, indicating that the higher a person's score for wanted inclusion, the more likely he/she will select a cooperative response. Additionally, the results suggest that the lower a person's score for wanted control, the more likely he/she will select a cooperative response.

REFERENCES

Hines, William W. and Douglas C. Montgomery, Probability and Statistics in Engineering and Management Science, 2nd ed., New York: John Wiley and Sons, Inc., 1980, p. 599.

Linden, Robert B., Personal Communication, December, 1981.

Rapoport, Anatol, Fights, Games, and Debates, Ann Arbor: The University of Michigan Press, 1960, p. 174.

Shaw, Marvin E. and Philip R. Costanzo, Theories of Social Psychology, New York: McGraw-Hill, Inc., 1970, pp. 255-259.

Whalen, Eugene, Personal Communication, December, 1981.

Effects of "Hyper" Descriptor and Cadets Second Order Self Perception On Their Ratings of A Hypothetical Cadet

David B. Porter and June I. R. Babson
U. S. Air Force Academy CO 80840

ABSTRACT

One hundred three USAFA cadets completed a person perception exercise, similar to Asch's "warm - cold" experiment. They formed an impression of a hypothetical cadet based on seven adjectives, including the descriptor "hyper" or "mellow". The descriptors "hyper" and "mellow" are common jargon in use at the USAF Academy, characterizing in general, one's attitude towards military training. The cadet raters also indicated whether they felt they were perceived by lower classmen as being "hyper" or "mellow". Results indicated that the hyper-mellow descriptors strongly influenced attributions concerning other traits. The results are discussed in relation to interpersonal perception on the part of cadets, faculty, and administration.

INTRODUCTION

Person perception, the human process of obtaining and organizing information about others, is one of the primary areas of study in social psychology. Most contemporary texts identify three important determinants of perception: the perceived, the perceiver, and the situation in which the perception takes place (Massarik and Wechsler 1959; Mitchell 1978). During interpersonal interaction, one's perception of another is partly influenced by both the real and imagined characteristics of the individual with whom one is interacting. The salient characteristics of the "perceived" range from physical appearance to observed behavior to inferences made by the perceiver of the perceived person's underlying motivation, values and personality. Although the relative importance of each of these characteristics depends on the situation, it was recognized as early as 1920 that the perception of separate individual characteristics influence the perception of other attributes and characteristics (Zalkind and Costello, 1962). Identified as the "halo effect", this process by which a general impression (which is favorable or unfavorable) influences the evaluation of other traits, has been shown to screen the accurate perception of many characteristics. Asch (1946) reported the results of an experiment designed to measure the halo effect of two dichotomous descriptors: "warm" and "cold". He found that when the word "warm" was added to five other words describing a hypothetical person, that person was also evaluated as being wise, humorous, popular and imaginative. By changing "warm" to "cold" in the initial description a radically different and much more negative impression was formed. Kelly (1950) showed that the "halo effect" gained strength when individuals were given the opportunity to interact with the individual about whom they had previously formed an impression. Asch (1946) also found that the effect was not nearly as great when more specific and less fundamental descriptors such as "polite" and "blunt" were used in place of "warm"

and "cold".

Several variables associated with the perceiver have also been identified which effect perception. Zalkind and Costello (1962) suggest that individuals who accurately perceive themselves and view themselves optimistically are more likely to accurately perceive others as well as to evaluate others more positively. Mitchell (1978) suggested that "people who are more secure, independent, and have a high tolerance for ambiguity are more accepting of others who are different from themselves," (p97). Massarik and Wechsler (1959) hypothesized different levels of self perception. While first order self perception is defined as how an individual views himself, second order self perception (how an individual thinks he is perceived by others) may be an equally important influence on his or her behavior.

The first author developed a classroom exercise to demonstrate these findings to his students. Although Asch's (1946) original experiment was generally replicated, the terms "hyper" and "mellow" were selected from the cadet vernacular to replace "warm" and "cold". An additional alteration was made by having participants indicate their second order self perception. This study had three primary purposes: 1) to examine the effect of the dichotomous descriptors "hyper" and "mellow" on the favorability of cadet's ratings of a hypothetical cadet on sixteen other dichotomous characteristics, 2) examine the effect of the cadet raters second order self perception (either "hyper" or "mellow") on the same ratings, and 3) to examine the differences in the ratings by the resultant four groups ("hypers" rating "mellows", "hypers" rating "hypers", "mellows" rating "mellows" and "mellows" rating "hypers").

METHOD

Subjects

The "hyper-mellow" exercise was administered to each of 103 third class (sophomore) cadets at the U.S. Air Force Academy.

Procedure

The exercise consisted of three parts. In the first part a description of a hypothetical cadet was presented. The exercise described the cadet as being: "intelligent, skillful, industrious, determined, practical and cautious." For approximately half (48%) of the cadets the fourth descriptor was "mellow" and for the rest of the cadets the fourth descriptor was "hyper." Cadets were given a few minutes to "form an impression" of this hypothetical cadet. In the second part of the exercise, cadets were asked to select an additional term from each of sixteen pairs of dichotomous descriptors. (Pairs are listed in Figures 1 through 3.) When all cadets had completed this part of the exercise and before any discussion of the exercise, the administrator (instructor) asked the following question: "If I (the instructor) were to ask the fourthclassmen in your element (immediate subordinates) to put you in one of two categories, hyper or mellow, which would it be?" Cadets were asked to write their response at the bottom of the exercise sheet. These sheets were then collected and the exercise was discussed in the context of person perception, stereotyping, and halo effect.

Scoring and Analysis

Three separate analyses were conducted on the resulting data. First the exercise sheets were divided into two groups: those in which the hypothetical cadet was described as being "hyper" and those described as being "mellow." The percentage of positive descriptors was tabulated for each of the sixteen pairs of descriptors and plotted in Figure 1. Individual responses to all sixteen pairs were then scored by assigning one point for each positive descriptor selected. The two groups were then compared for significant differences using a t-test. Second, response sheets were divided by second order self perception (as indicated by cadet responses to part three of the exercise) and were analyzed by the same method previously described. Finally responses were separated into four groups: hypers rating hypers, hypers rating mellows, mellows rating hypers and mellows rating mellows. Means, and standard deviations were computed and t-tests were used to identify significant differences among the four groups.

RESULTS AND DISCUSSION

Results of the first analysis are reflected in Figure 1. The sixteen negative descriptors are listed on the left axis and their positive counter-parts are listed on the right axis. Descriptors were sorted sequentially, the highest combined percentage of positive ratings listed at the top and those that received the lowest combined percentage of ratings listed at the bottom. Over 90% of the hypothetical cadets were rated as reliable, serious, honest and strong. Perhaps this is because these positive descriptors are more logically related to the six other initial descriptors (intelligent, skillful, industrious, determined, practical and cautious) than they are with the hyper-mellow variable descriptor. Conversely, the extremely low percentage of cadets selecting the "talkative" descriptor may reflect the perceived similarity of the initial descriptor "cautious" and "talkative"'s alternative "restrained". It is also interesting to note the low percentage of cadets rated as generous or unselfish, although neither of these terms seem to be logically related to the initial six descriptors. The relatively large percentage of cadets who rated the hypothetical cadet as being ungenerous (44%) and self centered (57%) may reflect the general image cadets hold of each other due to a competitive environment (Wood and Pacheco, Note 1; Porter, Note 2).

In the first analysis the independent variable was the description of the hypothetical cadet as being either "hyper" or "mellow". Figure 1 reflects that if the cadet was described as being "mellow" he was rated much more favorably than if he was described as being "hyper". If the hypothetical cadet was described as being "mellow" he was rated by over 80% of the cadets as being good-looking, humane, happy, popular, sociable, good natured, humorous, wise, and generous. The only overall negative descriptor of "mellow" was of being "restrained" and this is probably due to the initial descriptor "cautious". In contrast over 50% of the cadets described as being "hyper" were also described as being irritable, humorless, shrewd, ruthless, self-centered, unpopular, unhappy, ungenerous,

hard-headed and unsociable. These results suggest that the mellow-hyper descriptors is as strong an influence on perception as Asch's original warm-cold descriptors. An alternative analysis of the same groups was done by considering the total responses by individual rater rather than by separate descriptors. Again the results indicate that the "mellow" descriptor had a much more positive impact than the "hyper" descriptor. The 49 respondents, who rated a hypothetical cadet described as being "mellow", used more positive ratings ($X=13.50$, $SD=2.32$) than did those rating a hypothetical cadet described as being "hyper" ($X=8.75$, $SD=4.22$). A t-test showed that the differences in these two groups were significant at the .01 level ($t=7.174$, $df=102$) and substantiate the subjective comments made earlier concerning Figure 1.

In the second analysis responses were separated by the raters second order self perception. Seventy-four of the one-hundred-three subjects reported that their subordinates would categorize them as being "mellow". (It is interesting to note here, that an informal poll of freshmen cadets enrolled in Introductory Psychology indicated that these cadets would have only put about one third of the third classmen in their squadrons in the "mellow" category.) The remaining twenty-nine cadets indicated that their subordinates would have put them in the "hyper" category. Figure 2 shows the percentage of positive responses by rater of the descriptor pairs. The plot suggests that those cadets whose second order self perception was "hyper" rated the hypothetical mellow and hyper cadet slightly more positively than those whose second order self perception was "mellow". Although these differences were much less pronounced than those in Figure 1, "hypers" rated the hypothetical cadet more positively on fifteen of the sixteen dichotomous descriptors. The largest differences seem to be in the descriptor pairs sociable/unsociable (21% difference) and wise/shrewd (28% difference). These results are partially explained by the fact that a slightly higher percentage of cadets whose second order self perception was "hyper" rated hypothetical cadets who were mellow (55%). Cadets whose second order self perception was "mellow" rated 45% hypothetical mellow cadets, and 55% hypothetical hyper cadets. However, even taking this disparity into account (see Appendix A for complete derivation) cadets whose second order self perception was "mellow" rated the hypothetical cadet less positively than subjects whose second order self perception was hyper.

Although not statistically significant, these results are noteworthy. Even considering the different samples "mellow" cadets used fewer positive descriptors than predicted and "hyper" cadets used more positive descriptors than predicted. Two possible inconsistencies could contribute to these unexpected results. First, the stereotype of "hyper" cadets as being more irritable, humorless, shrewd, ruthless, self-centered, unpopular, and ungenerous than "mellow" cadets may be totally inaccurate. This is consistent with Allport's (1954) finding that some widely held stereotypes do not contain even "a kernel of truth." Another possible explanation deals with the accuracy of self-reported second order self-perceptions. Although it was not possible to identify inconsistencies on a case by case basis, a disparity in the percentage of mellow third class cadets as reported by second order self-perception (72%) and the results of the informal poll of their subordinates (33%) is apparent. This suggests

that the group of cadets who identified themselves as being "mellow" on the exercise were really made up of two separate and distinct groups: those who accurately perceived their subordinates' perceptions ("true mellow") and those who did not ("pseudo-mellows"). If this assumption is correct, we can interpret the combined less positive ratings of the mellow group to be the average of two significantly different groups. We would expect the "true-mellows" to use at least as many or more positive ratings as the "hypers" and the "pseudo-mellows" to use significantly fewer positive ratings. Because individuals in the pseudo-mellow category had either inaccurately assessed others' perception or could not accept others' perception, we would also expect them to perceive others less accurately and to be less tolerant of those who were different (Zalkind and Costello 1962, Mitchell 1978).

Finally responses were separated into four groups, plotted and analyzed. Figure 3 reflects the profiles of the four groups. There appeared to be a great deal of agreement between both "hyper" and "mellow" cadets rating hypothetical cadets described as being mellow. This is supported by comparing the average number of positive ratings by individuals in the two groups shown in Table 1. Although mellow cadets rated other mellow cadets slightly more positively ($X=13.6$, $SD=2.4$) than did hyper cadets ($X=13.4$, $SD=2.8$), the differences were not significant ($t=.25$, $df=47$, n.s.). This suggests that there is general agreement among all cadets that mellow cadets have many positive characteristics. In contrast both hyper and mellow cadets rated cadets described as being hyper less positively. The difference was highly significant in the mellow group of raters ($t=6.88$, $df=72$, $p<.01$) but only marginally significant in the hyper group ($t=1.78$, $df=27$, $p=.10$). This suggests that the group of cadets who identified themselves as mellow, or some relatively large subgroup of that group (i.e. "pseudo mellow"), maintained extremely negative as well as inaccurate stereotypes of other cadets described as being hyper.

CONCLUSION

Several important conclusions as well as some interesting conjecture can be drawn from the results of this exercise. The importance of these results is increased when one considers the essential role that accurate person perception has in leadership development which is the Academy's primary purpose. As Massarik & Wechsler (1959) pointed out over twenty years ago

"As an executive (leader) faces the myriad decisions he needs to make, it becomes quite clear that he must master two tasks: he must learn to see accurately the human, as well as the inanimate factors of the total scene; and he must acquire the skills of action which, while based upon accurate perception, tap the well-springs of behavior that ultimately lead to the successful attainment of personal and organizational goals (p46)."

The results of this exercise indicate that the halo effect associated

with "hyper" and "mellow", two commonly used descriptors in the cadet vernacular, may be a significant impediment to accurate interpersonal perception. This effect appears to be extremely severe in the one group of cadets hypothesized to be in the "pseudo mellow" category.

We also suspect that many members of the Academy faculty and administration are unfamiliar with the positive characteristics associated with the term "mellow" and the extremely negative images aroused by the term "hyper". We know of several occasions where cadets have been counseled for being "too mellow" by officers. It is little wonder that cadets frequently do not respond as expected to such counseling.

The results of this exercise also generally support and expand the conclusions reached in studies which have suggested the negative impact of excessive interpersonal competition on effort and performance (Wood and Pacheco, Note 1; Porter, Note 2). The overall negative attributions made toward "hyper" cadets may reflect the group's development of informal norms to suppress effort within the wing. Because the total number of rewards is generally fixed and distribution of these rewards is determined primarily through interpersonal competition, the general suppression of effort emerges as a significant superordinate goal for cadets (i.e. if everyone puts forth less effort and therefore incurs fewer personal costs then everyone improves their cost benefit ratio because the number of institutional rewards remained fixed). One way of enforcing this norm is by labeling those who put forth extra effort "hyper". The plethora of negative descriptors associated with this term suggests a variety of sanctions taken against "hyper" cadets by their classmates. For example, because "hyper" cadets are stereotyped as being unpopular and unsociable they are often not included in informal weekend social events; because "hyper" cadets are seen as ruthless, ungenerous and self-centered, other cadets feel justified in treating them unfairly and deceptively; and, because "hyper" cadets are seen as being irritable, humorless, and hard headed, their contributions and inputs are generally ignored. For this reason many cadets avoid being put in positions where they are expected to put forth extra effort (particularly effort directed toward enforcing military regulations) and many of those who feel compelled to perform in these positions inaccurately perceive or deny their image as "hypers" (i.e., pseudo mellows). The areas of person perception, halo effect and stereotyping are critical to the development of leadership. The results of this study strongly suggest a need for greater awareness of these factors as well as a need for further study.

Appendix A

Predicted mean positive ratings were computed for each group by multiplying the number of hyper rating hyper cadets in the sample times the average rating of hypothetical hyper cadets (8.75) and adding this to the number of hyper rating mellow cadets in the sample times the average rating of mellow hypothetical cadets (13.50) and dividing by the total sample size. Thus the predicted number of positive ratings are as follows:

$$\text{hypers } (13 \times 8.75) + (16 \times 13.50) / 29 = 11.37$$

$$\text{mellows } (41 \times 8.75) + (33 \times 13.50) / 74 = 10.86$$

The observed number of positive ratings for hypers were 12.27 (SD=3.71) and for mellows were 10.50 (SD=4.35).

EFFECT OF HYPER/MELLOW INITIAL DESCRIPTOR

ON PERCENTAGE OF POSITIVE RATINGS BY ADDITIONAL DICHOTOMOUS DESCRIPTORS

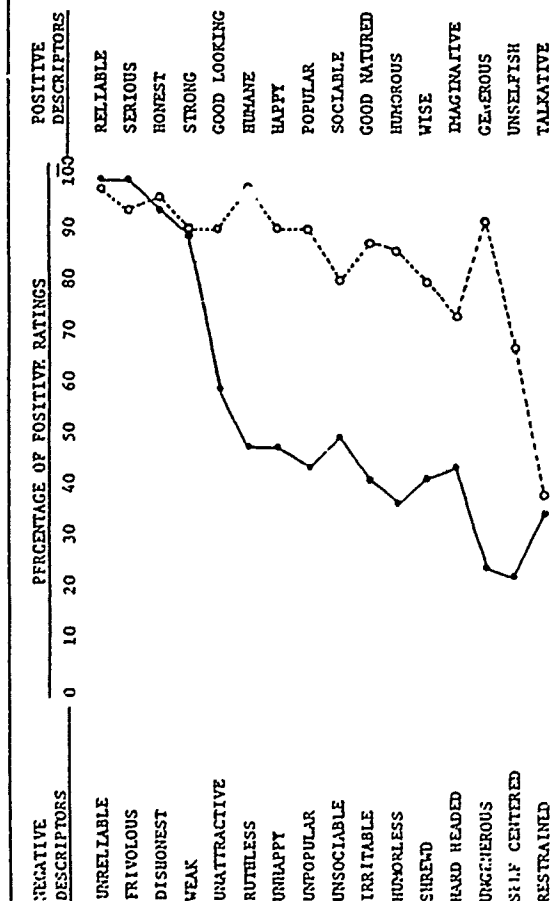


Figure 1

164

EFFECT OF RATER'S SECOND ORDER SELF PERCEPTION

ON PERCENTAGE OF POSITIVE RATINGS BY ADDITIONAL DICHOTOMOUS DESCRIPTORS

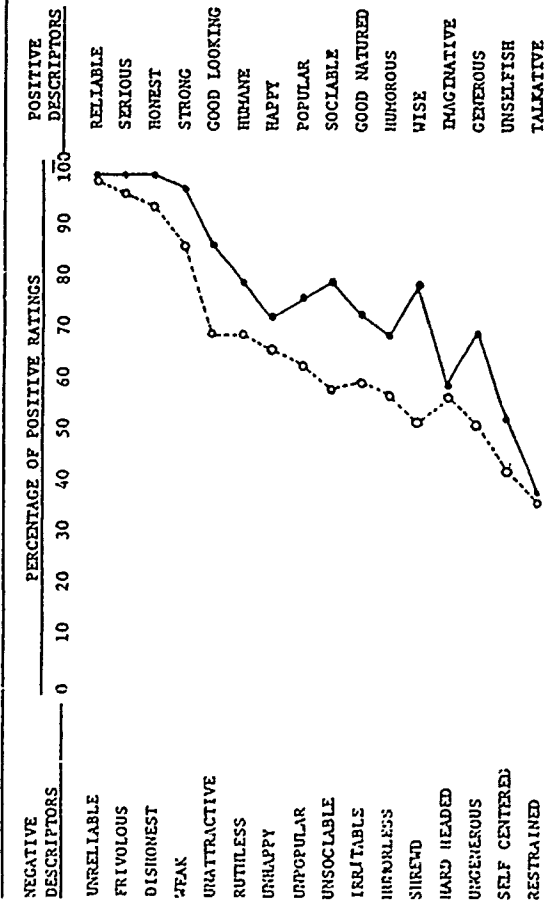


Figure 2

EFFECT OF GROUP ON PERCENTAGE OF POSITIVE RATINGS OF ADDITIONAL DICHOTOMOUS DESCRIPTORS

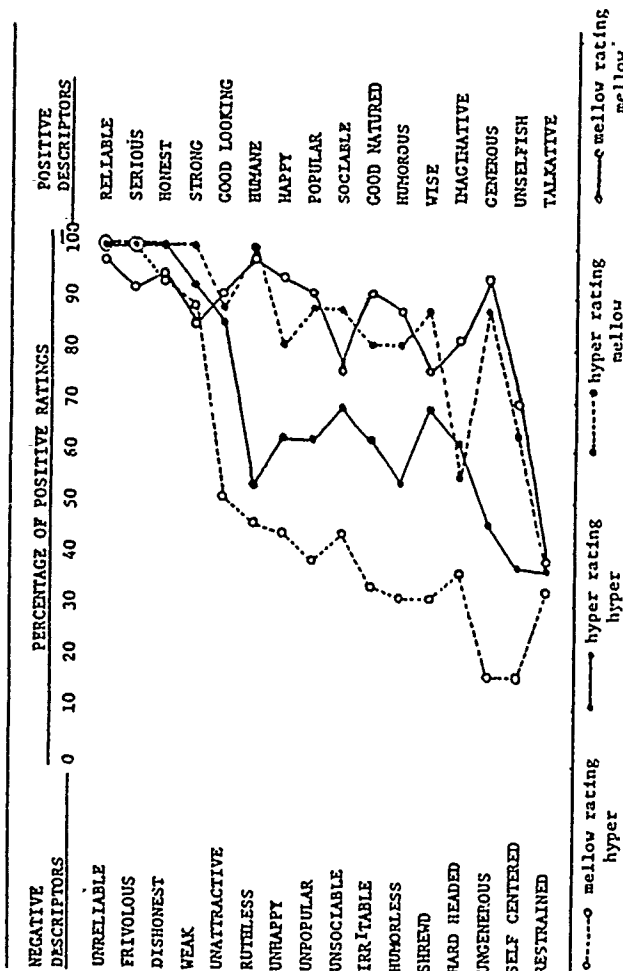


Figure 3

SAMPLE NUMBER, MEAN NUMBER OF POSITIVE DESCRIPTORS AND STANDARD DEVIATIONS BY GROUP

RATER'S SELF PERCEPTION		INITIAL DESCRIPTOR	
HYPER		HYPER	MELLOW
n	13		16
\bar{X}	11.0		13.4
SD	4.2		2.8
MELLOW			
n	41		33
\bar{X}	8.1		13.6
SD	4.0		2.4

Table 1

REFERENCE NOTES

1. Wood, F. R. and Pacheco, S. Effects of competition at USAFA (USAFA-TR-1981). USAFA, CO: Academy Technical Report, Publication pending, 1981.
2. Porter, D. B. Interdependent grading system at USAF Academy (USAFA-TR-1981) USAFA, CO: Academy Technical Report, Publication pending 1981.

REFERENCES

- Allport, G. W. The nature of prejudice. Boston: Addison-Wesley, 1954.
- Allport, G. W. Pattern and growth in personality. New York: Holt, Rinehart and Winston, 1961.
- Asch, S. Forming impressions of persons. Journal of Abnormal and Social Psychology, 1946, 60, 258-290.
- Hamachek, D. E. Encounters with the self (2nd ed). New York: Holt, Rinehart and Winston, 1978.
- Hastorf, A. H., Schneider, D. J. and Polefka, J. Person Perception. Reading, Mass.: Addison-Wesley 1970.
- Kelley, H. H. The warm-cold variable in first impressions of persons. Journal of Personality, 1950, 18 431-439.
- Kelley, H. H. and Thibaut, J. W. Interpersonal relations. New York; Wiley-Interscience, 1978.
- Massarik, F. and Wechsler, I. R. Empathy revisited: the process of understanding people. California Management Review, 1959, 1(2) 36-46.
- Mitchell, T. R. People in organizations: understanding their behavior. New York: McGraw-Hill, 1978.
- Wolman, B. B. Dictionary of behavioral sciences. New York: Van Nostrand Reinhold, 1973.
- Zalkind, S. S. and Costello, T. W. Perception: implications for administration. Administrative Science Quarterly, September 1962, 7, 218-235.

Nonverbal Communication in Military Counseling

Gary A. Packard, Jr.

and

Mickey R. Dansby*

Department of Behavioral Sciences and Leadership
USAF Academy, CO 80840

Abstract

An experiment was conducted to investigate the effects of nonverbal communication in a military counseling setting. It was found that potential officers (cadets) rated an officer counselor as more effective if he projected positive rather than negative nonverbal cues [$F(1,48) = 15.6, p < .01$]. Furthermore the results suggested that formal instruction increases awareness of nonverbal behavior. The presence of verbal content along with nonverbal cues did not significantly affect subjects' ratings of the counselor's effectiveness.

Nonverbal communication has been shown to play a significant role in the communication process (Knapp, 1972; Archer & Akert, 1977; Rosenthal, Archer, DiMatteo, Hall-Koivumaki, & Rogers, 1979). Certainly, one very significant communication situation is counseling on personal problems. As Benjamin (1981, p. 46) points out, in a counseling setting, "We hear with our ears, but we listen with our eyes and heart and skin and guts as well." It is evident that nonverbal communication is an essential part of the counseling process.

Military officers are frequently called upon to counsel subordinates on a number of topics ranging from career progression to personal finances and marital problems. Although many nonverbal cues relating to status difference (i.e., officer/enlisted, rank, etc.) are quite explicit in such military settings, many other cues are a product of the counselor's behavior. It would certainly be to the military officer's advantage to be able to communicate his concern via nonverbal channels, since much, if not most, of the message is conveyed via nonverbal means (Mehrabian & Weiner, 1967). Furthermore, the degree to which the client is sensitive to the nonverbal behavior of the counselor could significantly affect the results of the counseling session.

In the present study, we were concerned with seeing whether potential military officers (cadets) at the U.S. Air Force Academy are sensitive to the counselor's nonverbal cues in a structured, military counseling session between a superior and a subordinate. A second question of interest was whether senior cadets, who have received formal instruction in nonverbal communication, are more sensitive to nonverbal cues than freshmen, who have not received formal instruction. Finally, we also wondered whether the presence of verbal as well as nonverbal content would affect cadets' ratings of the counselor's effectiveness.

Method

In order to investigate the questions above, a 2 X 2 X 2 completely randomized factorial experiment was conducted in which the factors were nonverbal cues (positive or negative), class (freshman or senior), and sound (present or absent). It was hypothesized that positive nonverbal cues by the counselor would lead to more favorable evaluations of the counselor's effectiveness, that seniors would be more sensitive to nonverbal cues than freshmen, and that the presence of sound (i.e., verbal content) would affect ratings of the counselor's effectiveness little, if at all.

Subjects were 56 cadet volunteers, half of them seniors and half of them freshmen. There were eight females, making the sample roughly proportional to the female population at the Air Force Academy. Subjects were randomly assigned to nonverbal cue and sound conditions within class.

The stimulus materials were two 3-minute videotapes of simulated counselor-client interactions, identical except for the nonverbal cues present. In one tape, the counselor used positive nonverbal cues (i.e., feet firmly planted on the floor, slightly apart; arms open, not crossed; slight lean toward the client; frequent eye contact; occasional nods of the head; and hand gestures to emphasize speaking). In the other tape, negative nonverbal cues were used (feet crossed or one leg crossed over the other, slight slouching posture, infrequent eye contact, arms folded in lap, and few hand gestures). In both tapes the actors were the same: an officer counselor and a cadet client. The counselor was, in fact, a full time counselor at the Cadet Counseling Center. The client was a senior cadet with considerable acting experience. The script was the same for both tapes, and dealt with the cadet's decision to remain at the Academy or resign, a dilemma familiar to most of the subjects and a normal topic for a counseling session. In both tapes, the counselor wore an officer's uniform and the client wore a cadet's uniform. The camera angle (behind the client, looking over his shoulder to the counselor) and set (a pleasant office with no barrier between the counselor and client) were also identical for both tapes. The sound/no sound manipulation was accomplished by turning the volume of the TV monitor up or down, as appropriate.

Subjects viewed the appropriate tape either alone or in small groups, at the convenience of the subjects and the experimenter. Viewing sessions were scheduled randomly according to combinations of the independent variables. The tapes were shown on a standard videotape player and TV monitor. After the subjects entered the viewing room (a classroom) and sat down, the experimenter asked them to complete a standard consent form. Pre-recorded instructions were played and any questions were clarified. Subjects were shown the tape with the sound either on or off, according to experimental condition. After they viewed the tape, subjects were asked to complete a questionnaire on the counselor's effectiveness. The questionnaire consisted of 10 items on a semantic differential scale (helpful/not helpful, inattentive/attentive, empathetic/not empathetic, etc.), three rating scales on the helpfulness of the counselor, and one open-ended question on what the counselor might have done to make the session more productive. Subjects were debriefed and questions were answered after they completed the questionnaire.

Results

Data were analyzed via ANOVA. Results on the semantic differential scale (Table 1) revealed a significant main effect for nonverbal cues. None of the other main effects or interactions was statistically significant.

Table 1
ANOVA for Semantic Differential Rating Scale

Source	df	MS	F
Class (C)	1	311.1	1.99
Nonverb (N)	1	2444.6	15.61*
Sound (S)	1	380.6	2.43
C x N	1	16.1	0.10
C x S	1	5.8	0.04
N x S	1	330.3	2.10
C x N x S	1	7.1	0.05
Error	48	156.6	

*p < .01

Examination of cell means in the nonverbal cue conditions revealed that, as predicted, positive nonverbal cues ($M=24.1$; minimum score=10, maximum=70; lower score indicates more favorable rating) led to more favorable ratings of the counselor's effectiveness than did negative nonverbal cues ($M=37.3$).

Analysis of the other questionnaire items showed a similar trend. There was a tendency to be more confident that the client would return to confer with the counselor if the nonverbal cues were positive ($M=2.75$; minimum score=1, maximum=7; lower score indicates more favorable rating) rather than negative ($M=3.50$). Subjects also indicated that they would be more comfortable seeing the counselor if the nonverbal cues were positive ($M_{pos}=3.61$, $M_{neg}=4.64$) and that they thought the counselor was more helpful when positive cues were present ($M_{pos}=3.18$, $M_{neg}=4.11$). The ANOVA revealed that the difference in ratings on the probability of a return visit was statistically significant [$F(1,48)=7.75, p<.01$]. The rating differences in the other two variables, while showing the same general trend, were not statistically significant at conventional levels ($p=.075$ and $p=.067$, respectively). None of the other effects was statistically significant for these three variables.

Responses to the open-ended question on ways that the counselor might make the session more productive revealed a trend for seniors to make specific suggestions for improving nonverbal behavior (e.g., leaning forward, nodding head, etc.), while freshmen did not (seniors made 11 specific suggestions, freshmen only one; $\chi^2(1)=9.8, p<.01$).

Discussion

The results support the hypothesis that cadets are sensitive to nonverbal cues in the military counseling setting. Clearly, positive nonverbal cues were interpreted more favorably than negative nonverbal cues. This finding underscores the importance of military counselors being aware of the effects of their nonverbal behavior if they wish to be more effective.

Interestingly, senior cadets and freshmen cadets did not differ in their ratings of the counselor's effectiveness. Originally, we thought that seniors might be more sensitive to nonverbal behavior because they had received academic instruction on nonverbal communication, while freshmen had not. Even though seniors' ratings of the counselor's effectiveness were equivalent to those of the freshmen, the seniors were able to describe specific ways in which the counselor might improve his nonverbal behavior. This suggests that the formal instruction on nonverbal communication (taken during the junior year) may have made the seniors more aware of nonverbal cues. Perhaps the affective response to nonverbal behavior is well-ingrained through years of experience with nonverbal interactions with others, whereas the cognitive response to these nonverbal cues is facilitated by formal instruction. Freshmen knew something was wrong but were not able to say what; seniors not only knew something was wrong but were able to identify nonverbal communication as the problem.

Further evidence that formal instruction may increase sensitivity to nonverbal behavior comes from a comparison of responses from seniors in a counseling class ($n=5$) to those of the randomly selected seniors in the positive nonverbal, no sound group ($n=7$) of our experiment. The counseling class members were asked to view the videotape and rate the counselor's effectiveness in the same way as subjects in the main experiment. The seniors from the class rated the counselor's effectiveness more negatively on the semantic differential scale ($M=26.6$, lower score = more favorable response) than did the seniors selected at random ($M=16.6$). Comparison of the variances for the two groups showed that they were significantly different [$F(4,6)=10.29$; $p<.05$]. A t -test using separate variance estimates also revealed a significant difference between mean ratings [$t(4.56)=1.88$, $p<.10$]. Of course, since the seniors had self-selected

to take the class, it is possible that factors other than their formal instruction (e.g., greater interest or aptitude for counseling) contributed to the difference in ratings. Nevertheless, the results are consonant with the hypothesis that formal instruction increases awareness of nonverbal cues.

As expected, there was no significant difference in the ratings of counselor effectiveness when verbal cues were present in addition to the nonverbal cues. This agrees with findings by Mehrabian and Weiner (1967) that verbal content may be less important in the total communication package than nonverbal content. In essence, the reinforcement by verbal content adds little to what is communicated by nonverbal cues.

The present experiment highlights the importance of nonverbal communication in a military counseling setting. Clients are sensitive to the nonverbal behavior of the counselor and interpret nonverbal cues as a significant part of the counseling process. We would do well in the military services to pay attention to those nonverbal cues under the control of the counselor if we wish to increase counselor effectiveness. Perhaps one way of increasing awareness of nonverbal behavior is, as suggested in the present study, through formal instruction. Further research might clarify the relative importance of formal instruction in helping potential military officers understand and pay attention to their own and others' nonverbal communication.

References

- Archer, D., & Akert, R. M. Words and everything else: Verbal and nonverbal cues in social interpretation. Journal of Personality and Social Psychology, 1977, 35, 443-449.
- Benjamin, A. The helping interview. Boston: Houghton Mifflin, 1981.
- Knapp, M. L. Nonverbal communication in human interaction. New York: Holt, Rinehart, & Winston, 1972.
- Mehrabian, A., & Weiner, M. Decoding inconsistent communications. Journal of Personality and Social Psychology, 1967, 6, 109-114.
- Rosenthal, R., Archer, D., DiMatteo, M., Hall-Koivumaki, J., & Rogers, P. Body talk and tone of voice: The language without words. Psychology Today, 1974, September, 64-68.

STOPOUT RETURNEES TO USAFA

Jim Clifford
USAF OMC, Randolph AFB, TX

Abstract

Stopout is the name of a new program at the United States Air Force Academy, whereby cadets who have completed two years of study at the Academy are granted the option of taking a leave of absence for up to a year. The first group of Stopouts returned in the summer of 1981. This study is composed of interviews conducted with twenty of these returnees to investigate their experiences during Stopout. The information obtained was overwhelmingly positive, in addition to providing valuable information to United States Air Force Academy future Stopouts, and future returnees. The study also provides direction for future work and improvement on Stopout and related programs.

The Problem

Stopout is the official name of a program here at USAFA whereby a cadet can choose to take leave for up to a full year, and be guaranteed reinstatement in the Cadet Wing if he/she desires. They are then allowed to continue the academy program one year behind their original class. Prior to this, the four years at USAFA were a "lock step" cycle for virtually all cadets. A provision such as Stopout was practically unheard of except in extremely rare situations. For example, a cadet who was doing very well, but then experienced some medical problem which required a semester of recovery, might be granted a leave for this period. This is officially termed medical turnback. The only other circumstance where such a leave might be granted is known as administrative turnback. Here, too, a cadet applying for such a leave of absence must demonstrate that he/she is performing well as a cadet. They must also indicate a severe hardship with their family which requires their presence at home. Very few of these turnbacks are granted (one or two per year). Thus, there were leave provisions for medical or family related circumstances, but none whatsoever for personal reasons of the individual cadet. However, under the Stopout program, cadets can now leave, if they wish, for a period of up to 13 months, commencing right after the completion of spring semester, and ending in the summer a year later. They must have completed their sophomore year at USAFA, and not have any academic or military deficiencies. This program does not currently exist at any other military academy.

I was among this first group of Stopouts, which explains in part my interest in the people who, like myself, have returned.

The subjects of this study were the group of twenty cadets who returned to the Cadet Wing from the first Stopout program. This study does not include any medical or administrative turnback cases.

In the study I have attempted to find experiences and perceptions common to Stopout returnees. I have also attempted to obtain ideas and information which can help the future of the Stopout program in general.

Review of the Literature

Although many civilian institutions of higher learning have existing programs similar to Stopout, no comparable data is available. Many civilian students at the undergraduate level regularly leave, returning later to the academic environment. This frequent coming and going of students is viewed as a normal part of the maturational process. Hence, it is not practical nor logistically feasible to develop an official program to monitor the transitions of these students.

Dean K. Whitla, the Director of the Office of Institutional Research and Evaluation at Harvard, was asked to evaluate the curriculum and program at USAFA. According to those close to the decision, the letter Dean Whitla wrote to USAFA detailing his findings and recommendations was among the most influential information in the Academy Board's decision to initiate the Stopout program. Following is a summary of that letter delineating the results of his evaluation:

A consistent attrition rate of 40% at USAFA, while being comparable to most civilian colleges when at the four year point, is unacceptably high for a school like USAFA. It is unacceptable in terms of the lost tax dollars that are represented in that figure. It is also unacceptable when one considers the number of helping programs at work at USAFA, to keep cadets from dropping out, go far beyond anything found at almost every civilian school nationwide. Finally, this high attrition rate is unacceptable mostly because the great majority of these dropouts are our best people. Only "...a few are academically or physically marginal - the majority are your most attractive and able cadets - a serious loss to the Air Force and the nation."

He also believed the admissions area was the wrong place to concentrate effort in hopes of changing the dropout rate significantly. "Almost all improvements will, I believe, come from changes within the four years of cadet experience while they are here at USAFA." Thus, the emphasis now should be not on who we bring in, but on what we do to them while they are here.

Whitla contends that granting students a leave of absence is "the most important factor" in achieving lower attrition rates. He believes that USAFA should definitely be able to handle "deviations" from the four year lock-step program, and should institute a stopout policy. Further, he asserts that there are several reasons why a rigid policy of four consecutive years without interruption may be one of the worst strategies to follow for getting an education:

(1) There is NO theory to support the straight four year plan for education. (Piaget, White, Perry, Kohlberg, Loevinger, Holland, etc.). This research indicates that educational growth is "NOT a monotonic linear function". It is in fact ridiculous to suggest going to school for four straight years and expect a steady upward slope of learning.

(2) A four year education plan will be "characterized by growth spurt periods, plateaus, frequent regressions, and these vary markedly from student to student.

(3) In fact, these variations apply more to males than females. This fact has even greater consequences at USAFA, since the student population is predominantly male (approx. 85%).

(4) "Six months of work and six months of travel work wonders for students. After a leave of absence they return refreshed, much more mature, and ready to resume their studies in ways that are impressive." This refreshing, maturing, and remotivating effect should be of special interest to USAFA since (according to Dean Whittle) cadets tend to be less mature than their counterparts at civilian colleges. The reality of fatigue, which Stopout can help remedy, is explained by Dean Whittle. He states:

Cadet schedules are far too demanding; we tallied the hours demanded by course work, study, military and athletics. Cadets have virtually no discretionary time. In fact there is little time for sleeping; many even stand in class simply to remain awake. Highly motivated, energetic young adults can, for short periods of time, meet such a schedule, but to do so for a period of years cannot be productive...

There are many beneficial reasons for instituting and continuing a leave of absence (Stopout) program at USAFA. Virtually the only alternative to such a program available to cadets is to simply quit/resign outright. Under these circumstances, the Academy "forgets you", and no longer considers you connected with USAFA in any way. If this former cadet then wishes to reapply, he/she must initiate a very lengthy and discouraging process with no assistance from the Academy. In short, without Stopout, no effort would be made to save these valuable human resources. Tax money invested in these individuals' prior schooling at USAFA would also be wasted.

FINDINGS

The positive aspects of Stopout. Half of the returnees mentioned that Stopout enabled them to see the positive aspects of USAFA more clearly. They could see the opportunities, the experiences, the education, the greater challenge, the security, and the economic advantages. In addition to making the positive aspects of USAFA clearer to them, Stopout also faded the intensity of the negative things which were involved in their decision to leave. In short, they stated, "The longer you are away from USAFA, the better it looks. This becomes important in your decision to return."

Half the subjects also mentioned that Stopout really helped them mature. Several felt that this was because they were not so sheltered, stifled, and isolated on Stopout as they were at USAFA. They related that they had never experienced independence because they entered USAFA less than one month after graduating from high school.

A third of the subjects found that they gained a more realistic perspective of the "outside world". Many of them consequently decided that "the grass isn't all that green out there."

Half said that Stopout allowed them to "reevaluate whether they really wanted to be here or not." They were able to do this in "a more objective climate" than if they were at USAFA.

Eight of the twenty said that "the best aspect was taking a much needed break to rest from the constant grind, highpaced/pressure program at USAFA." This break helped them "charge back up" to perform more efficiently upon their return.

Eight of the twenty also said that the Stopout program is the one reason that they are still cadets right now. They were convinced that they "hated this place" and were definitely not coming back. "I was out of here - gone for good, and I know for sure I was never coming back. But, I changed my mind as a result of my experiences on Stopout, and decided to return. I was sure glad they had Stopout!" "I took Stopout as a joke - I laughed at their suggestion of the idea, but when I got towards the end of my year, I wasn't laughing any more. They were smarter than I was! I thought I would never change my attitude, but it was as though they knew that I would."

One fourth said that they benefitted from learning to relate to "normal and everyday people", instead of always working only within the singular environment and people of USAFA. They also said that as a result, they improved their human relations skills. Working with and being around this greater variety of people afforded these Stopouts another luxury they had not known as cadets. They were accepted as unique individuals, free from any stereotypes that people held about them as cadets.

A third of the subjects said that they didn't let all the little, "nit-picky things here" get to them as much anymore. They stated that they "wouldn't get all bent out of shape over all the Mickey Mouse things that go on here." They now tend to take bothersome or disappointing things more in stride.

Negative aspects of Stopout. Half of the subjects said that the one negative thing was being in a different class now. You aren't in the same class with all your old friends. You may also feel some resentment from certain members of your old class because you "wimped out" in their eyes. This may also happen with people in your new class when they realize that you already know some of the firstclass cadets personally, even if you don't let this compromise your job performance. Most found that this is something you must deal with, but that the significance of this problem was minimal after you worked through it and resolved it.

Perceived changes in perspective/attitudes. Almost all of the responses were identical to those reported for question two.

One idea mentioned by several people was that long term goals are now an important component of their attitude about being at USAFA. This attitude incorporates things beyond the scope of just the cadet life and includes a degree of commitment to a career in the Air Force, career plans, and objectives they want to accomplish after graduating from USAFA.

Concerning friends of the same sex, half of the subjects felt that their strongest friends are at USAFA. This is mainly because they believed that "having strong relationships is critical to survival here." This philosophy accounted for

the strong friends they acquired at USAFA. However, they considered the friends they encountered on Stopout to be "special" because they didn't become friends based on extreme circumstances such as those they experienced at USAFA. Rather, they felt they had more "choice" in the matter of making friends when they were outside of USAFA because you didn't "have to" make friends with anyone. Several people mentioned that because of the problems caused by being in a different class (mentioned previously), "you can quickly identify your true friends."

Why they returned. A third of the subjects said that friends at USAFA had a lot of influence on their decision to come back. As mentioned previously, this resulted both from talking with them and from the mere fact that they did have so many good friends at the Academy.

Half mentioned economic factors such as cost of education, and a poor future in the job market. They saw more security and sound economics when they reevaluated what USAFA had to offer as compared to the outside. Half also mentioned that their appraisal of the education here and its value to them was a factor in their decision to return.

A third of the subjects responded that their decision to return consisted of a simple analysis of their main goals in life. This analysis showed them that the best/most likely way of reaching those goals was by returning and completing the program at USAFA. Interestingly, only two people mentioned that wanting to be a pilot was the main thing that brought them to the conclusion that they should return and graduate. In fact, several mentioned that they had a more moderate attitude about the "need to be a pilot." They felt that this attitude was tempered by a realization they now have: "there is a lot of hype and propaganda here to push you into being a pilot. Its not as important to me anymore - I'll make sure I'm doing what I like most, and if that happens to include flying, fine. If not, that's fine too."

Information for Stopout returnees. One quarter mentioned that USAFA did not provide a system which would facilitate their transition back into Cadet Life. Several felt they were treated as nonentities. For example, some of them were not on the computer listing so they had no classes scheduled for the fall semester. Some of them said that their official communications, appointment slips, and grades never cam to them - they had to run over to their old squadron and pick these things up. Several said that they still aren't in the Wing Alpha Roster (complete listing of all the members of the Cadet Wing).

The Stopouts found it difficult to readjust to the very controlled lifestyle and the extreme academic load at USAFA. Almost everyone agreed that there was a vital need to design a program to help Stopout returnees assimilate back into the Academy. However, they warned against structuring such a program too extensively. They suggested building enough flexibility into the program to allow each person to have some input into how they as an individual might best adapt back into the Wing. They disagreed that "the best way to re-blue them is to make them serve as Cadre for Basic Training as soon as they get back." Rather, they felt that assimilation and adaptation back into USAFA could best be accomplished through a much more gradual process. They felt this was important because the faster they were forced to assimilate back into the Wing, the faster they got burned out again.

Another thing several returnees had to deal with was the fact that they had just spent over a year in an environment which not only tolerates, but often encourages dishonest practices. For example, cheating was frequently encountered as accepted practice at civilian universities by those Stopouts who attended school. They warn that returning to a system which does not accept these practices requires extra awareness initially.

Information for potential Stopouts. One line of thinking which almost every subject had heard countless times was, "I really think Stopout is a great idea and I'd love to do it, but I just know I wouldn't come back." First of all, this statement assumes that you will tend to become convinced on Stopout that USAFA is so terrible and you sure aren't going to put yourself right back in the midst of all that again. Yet, as one can see quite clearly from this report, exactly the opposite tends to happen. The trend seems to be that the longer one is away, the more clearly one sees the positive aspects of USAFA, and the less intense are the visions of all those so terribly negative aspects. Upon return, one finds that these negative things that used to be so important don't seem to matter as much anymore. In choosing not to leave simply because you might not come back, one ignores the trend that these returnees have demonstrated.

Secondly, this reasoning ignores the fact that it is not necessarily "bad" if you do indeed find something that you really want more than being at USAFA. "It is better for both you and the Air Force if you find out early that being here is not what you want. You can quite easily take advantage of such knowledge now, rather than waiting until you have been flying a \$20 million airplane for two years to discover that you'd really rather be doing something else." Again, one of the most positive aspects of Stopout for these people was that it "allowed them to analyze exactly what they really wanted to do most." The Stopouts agreed that you shouldn't turn down the program out of fear of not returning. They felt that to do so would be a "cop-out". "I hope they all realize that it took a lot more guts to leave than it did to stay."

The Interdependence of International Affairs: Social Psychology as a
Politico-Military Tool

Richard W. Bloom
Det 1, AFCOS
San Antonio, Texas 78243

Abstract

The international viability of the nation-state depends on interdependent dimensions. The military dimension of this viability itself subsumes interdependent components--including the social psychological. The use of this component to resolve a politico-military dilemma is illustrated through application of research on attitude change, attitude-behavior relationships, and mass communication effects. This example is one of many wherein DOD psychologists can aid the nation-state through assessment, intervention, and evaluation in an interdependent world.

The key concept in appraising the international viability of a nation-state is interdependence (Vaky, 1980)--economically by variations in international currency markets, trade and tariff policies, and energy price structures; socially by transmission of values, morals, and life styles through mass media; militarily by strategic fluctuations induced by coups, weapons transfers, and shifts in ideological allegiance. In fact, economic, social, military, and other dimensions of the nation-state are themselves interdependent as can be deduced from many political dilemmas (Maynes & Ullman, 1980). For example how do the South African government's policy of apartheid, its strategic minerals and geopolitical import, its anti-Communist stance, its formation of an unofficial platinum cartel with the Soviet Union, its denunciation by the United Nations General Assembly and the Organization of African Unity, its vital role in feeding and supplying many African nation-states, its concern for its own existence, its nuclear weapons capability, and its contribution to the United States economy through multinational corporations translate into a coherent Reagan policy towards Johannesburg? Interdependence is an essential factor in any consideration of a polity--the nation state--in the international arena.

As psychologists in the DOD, we deal primarily with the military dimension of a nation-state's international viability. This dimension is not only interdependent as described above, but also subsumes a host of interdependent components. These components include the operational, the logistical, the technological, and the social psychological (cf. Halperin, 1977).

This, last, the social psychological, has been explored throughout history by the elite of military theorists (e.g. Sun Tzu, 1963/C. 400 BC; Machiavelli, 1940/C. 1430; Hart, 1954). It stresses the interdependence of the non-military with the military. And it is a realm in which DOD psychologists can improve their nation-state's inter-national viability.

To illustrate, I shall describe a real-world situation and suggest a social psychological resolution.

The situation encompasses the strong anti-nuclear attitudes of many Western Europeans. These attitudes have been manifested by anti-nuclear mass demonstrations (generally with larger an -US than anti-USSR positions), scathing newspaper editorials and reporting (again more anti-US than anti-USSR), and governmental waffling on allowing the United

States to deploy GLCMs and Pershing-2s on their territory and politically supporting the US production of "neutron bombs" (cf. The stormy summit, 1981). These attitudes presage the maintenance of Soviet nuclear and conventional superiority in the NATO-Warsaw Pact arena, the further eroding of NATO cohesion and concurrence, and the "Finlandization" of Western Europe (cf. Reagan & Brezhnev, 1981).

To DOD psychologists it should be quite apparent that the underlying phenomena in this situation include attitude change, attitude-behavior relationships, and mass communication effects. Attitude change phenomena are reflected in governmental shifts from 1979 to 1981 on additional United States nuclear deployment and populace shifts from post World War II attributions of the Soviet Union as war monger to present ascription of this role to both superpowers or the United States alone. Mass communications effects include success of multi-source Soviet propaganda and the Leninist "useful idiot" modulation of collective political behavior. Perhaps most important is the nature of attitude-behavior relationships. It is axiomatic that concern is less with Western European attitudes than how these will behaviorally translate in international security affairs.

Through a knowledge of relevant social psychological research and a healthy respect for cross-cultural limitations, DOD psychologists could present accurate information resolving the situation as follows.

First, do we change Western European attitudes through persuasive communications or through inducing counterattitudinal behavior (Jaccard, 1981)? Given the practical difficulties of inducing counterattitudinal behavior in interrelated but distinct mass populations, we would choose the former--although the latter has had some successes (Reich, 1970/1933).

Second, will our information involve the central route to attitude change with attention to content (e.g. Tesser, 1978) or the peripheral route with attention to context and typically researched source variables (Petty & Cacioppo, 1981)? Or should we use some combination of the two approaches? Given that the implications of nuclear deployment have high personal relevance to politicized Western Europeans and that we seek an enduring attitudinal change supporting this deployment, we should be much more concerned with content than context (Burnkrant & Sawyer, in press; Cook & Flay, 1978). Further, context variables should only be of concern as far as they potentiate the known mechanisms by which content approaches are effective (Eagly and Himmelfarb, 1978).

The importance of known mechanism becomes a third issue. Given that content approaches to attitude change in matters of high personal relevance work primarily through situations wherein individuals increase cognitive processing to message content (Petty, Cacioppo, & Goldman, 1981), factors enhancing such processing should be included in our information--those inhibiting such processing should be excluded. Thus we should not use rhetorical questions, for they decrease cognitive processing to message content in situations of high personal relevance (Petty, Cacioppo, & Heesacker, 1981). There should be no "what price freedom" attempts.

So, too, the technique of an unexpected, seemingly incongruous source--content pairing should not be used (Wood & Eagly, 1981). Although it can often lead to more attitudinal change than expected pairings, it usually does so through situations with decreased cognitive processing of message content. In fact, in matters of high personal relevance, expected pairings are not only as effective or more effective than unexpected pairings, but also work through situations with increased cognitive processing of message content. Thus we need not tarry over finding leftist Social Democrats or scrub-cheeked "Greens" who embrace nuclear deployment with swelling hearts.

With attention and comprehension held constant, message repetition usually increases cognitive processing of message content as does the multiplication of the message's sources (Harkins & Petty, 1981). Factors that increase motivation to cognitively process message content and increase pre-existing cognitive schemas relevant to our information will also potentiate the effect of our efforts (McGuire, 1969; Wood & Eagly, 1981). Thus our sources should be perceived by the Western European populace as likely proponents or neutralists towards GLCM and Pershing-2 deployment. Our presented information would address how deployment would attenuate plausible threats to personal concerns such as the health and well-being of children, physical safety, and adequacy of food, clothing shelter, and fuel. The actual concerns used and their prioritization would be empirically delineated through appropriate assessment, eg., polling of salient concerns and recognition of what has proven most amenable to mass communications influence.

A fourth issue involves not the context or source factors of our information but the context into which our information shall be placed. This context encompasses competing messages with a myriad of persuasion and counterpersuasion attempts. We can work this context for us by noting that increasing cognitive processing (as described above) towards message content can induce resistance to counterpersuasion (Snyder and Wicklund, 1976). Whatmore, aspects of our information which increase the number of message evaluators or induce distraction towards competing information will decrease cognitive processing of the latter's message content (Petty, Cacioppo, and Heesacker, 1981). Finally, once attitude change occurs, group polarization effects would render counterpersuasive techniques less likely to succeed (Myers, 1980). Here we would be allowing the "social movement" aspects of Western European anti-nuclear attitudes to guard against any "reconversion" experiences (Toch, 1965). These aspects would include a need for absolutes and a closing of the mind. Note also that we are here addressing the vital interaction of mass communications and social structure--vital to the success of any change campaign (Roberts and Bachen, 1981).

Now that we have decided on a methodology of attitude change, how can we maximize the probability that this change will foster desired behaviors such as a decrease in anti-US components of political demonstrations, an increase in demonstrations against the Soviet Union, and, ultimately, GLCM and Pershing-2 deployment? Social psychological research suggests the following. First, by increasing the temporal stability of attitude change--a goal of our methodology--desired behaviors will more likely occur (Bentler and Speckart, 1981). Second, we can further the likelihood of occurrence by focusing on behaviors that are under the voluntary control

of the populaces in question (Bagozzi, 1981) and have at least a modicum of instrumental benefit. Thus a voluntary behavior such as decreasing political demonstrations against further nuclear deployment on Western European soil would be a more viable goal than a sudden termination of the development of Siberian-Western European pipeline. The latter goal, while voluntary, lacks immediate benefit to Western European interests--although it could be a later goal of a concerted attitude change program.

Third, we should provide ample, direct experience on which to base desired attitudes on both molar, e.g., diplomatic speeches and TV advertisements, and community, e.g., public meetings and cultural exchange programs, levels (Beltran, 1976).

Lastly, we must focus on all relevant components of the desired attitude (Bagozzi, 1981). The attitude should address actions, targets, and behavioral contexts congruent with those of the desired behavior. It should consist of both cognitive and affective factors that are mutually consistent and congruent with the desired behavior. And it should be accurately linked through path analysis before implementation of our plan with other variables implicated in attitude-behavior relationships--behavioral intention, habit, and proximal and distal behaviors. It should not just merely consist of a negative valence towards an objects. A global "the Soviets are bad guys" approach will not work.

Now that we have addressed a methodology of attitude change and its probability of generating desired behaviors, there are only a few caveats before our information can be selected and presented. First, the individual differences literature on persuasability is assumed to be irrelevant when modifying the attitudes of populaces, unless there are significant differences in the proportions of sex, cognitive, and other relevant variables within these populaces. Second, the relative efficacy of mass communication sources has not been discussed, because there is no significant research (See Roberts and Brachen, 1981). What is available pertains to patterns of usage and the uses/gratifications people derive from specific sources. A positive correlation between frequency of exposure and manifestation of the desired attitude and/or behavior is usually assumed. Third, the "process" literature on how attitude change and attitudinally induced behavior change occur is only salient in suggesting operative mechanisms generating the efficacy of techniques specified in specific situations. Analogous to psychotherapy research which also deals with attitude and behavior modification, there is no need to embrace myths of universality and uniformity (Kiesler, 1971) involving cognitive dissonance, balance, impression management, etc. Fourth, analogous to psychotherapy research we do not assume a uniformity of intent, etiology, and attitude behind the behavior of anti-nuclear demonstrations. Assessment through polling and through analysis of speeches and documents would discriminate separate attitudinal targets--each, perhaps, requiring a significant intervention. Lastly, and continuing with the analogy to psychotherapy, the press of life events will often not reduce the legitimacy of a presenting problem. Here, even with the advent of the Geneva talks on weapons limitations/reductions, the attitudes and behaviors of Western Europeans will still be crucial. For example, an obvious refusal of the West German government to accept deployment would obviate the need for serious Soviet negotiation.

There are many other examples of social psychology as a politico-military tool. These include path analysis for latent variables in

insurgency scenarios, delineation of cohesive and desintegrative factor: in political movements, and cognitive analysis of strategic thinking. For DOD psychologists the opportunity to aid the nation-state is immense in the 1980's and beyond.

REFERENCES

- Bagozzi, R. P. - Attitudes, intentions, and behavior A test of some key hypotheses. Journal of Personality and Social Psychology, 1981, 41, 607-627.
- Beltran, S. L. R. Alien premises, objects, and methods in Latin American communication research. Communications Research, 1976, 3, 107-134.
- Bentler, P. M. and Speckart, G. Attitudes "cause" behaviors: A structural equation analysis. Journal of Personality and Social Psychology, 1981, 40, 226-238.
- Burnkrant, R. E. and Sawyer, A. G. Effects of involvement and message content on information processing intensity. In R. Harris (Ed.), information processing research in advertising. Hillsdale, N. J.: Erlbaum, in press.
- Cook, T. D. and Flay, B. R. The temporal persistence of experimentally induced attitude change: An evaluative review. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 11), N. Y.: Academic Press, 1978.
- Eagly, A. H. and Himmelfarb, S. Attitudes and opinions. In M. Rosenzweig and L. Porter (Eds.), Annual Review of Psychology (Vol. 29) Palo Alto, Calif.: Annual Reviews, 1978.
- Halperin, M. H. The President and the military. In J. E. Endicott and R. W. Stafford, Jr. (Eds.), American defense policy, (4th ed.). Baltimore: Johns Hopkins University Press, 1977.
- Harkins, S. G., and Petty, R. E. Effects of source magnification of cognitive effort on attitudes: An information-processing view. Journal of Personality and Social Psychology, 1981, 40, 401-413.
- Hart, L. Strategy, N. Y.: Praeger, 1954.
- Jaccard, J., Toward theories of persuasion and belief change. Journal of Personality and Social Psychology, 1981, 40, 260-268.
- Kiesler, D. J. Experimental designs in psychotherapy research. In A. E. Bergin and S. L. Garfield (Eds.), Handbook of psychotherapy and behavior change, N. Y.: John Wiley and Sons, Inc. 1971.
- Machiavelli, N. The prince. The discourses. N. Y.: Modern Library, 1940.
- Maynes, C. W., and Ullman, R. H. Ten years of foreign policy. Foreign Policy, 1980, 40, 3-17.
- McGuire, W. J. The nature of attitudes and attitude change. In G. Lindzey and E. Aronson (Eds.), The handbook of social psychology (2nd Ed.) (Vol. 3), Reading, Mass: Addison-Wesley, 1969.
- Myers, D. G. Polarizing effects of social interaction. In H. Brandstatter, J. H. Davis, G. Stocker-Kreichgauer (Eds.), Contemporary problems in group decision Making. N. Y.: Academic, 1980.
- Petty, R. E. and Cacioppo, J. T. Attitudes and persuasion: Classic and contemporary approaches. Dubuque, Iowa: Wm C. Brown, 1981.
- Petty, R. E. Cacioppo, J. T., and Goldman, R. Personal involvement as a determinant of argument-based persuasion. Journal of Personality and Social Psychology, 1981, 41, 847-855.

Petty, R. E., Cacioppo, J. T., and Heesacker, M. Effects of rhetorical questions on persuasion: a cognitive response analysis. Journal of Personality and Social Psychology, 1981, 40, 432-440.

Reagan and Brezhnev. The New York Times, November 19, 1981, p. A31.

Reich, W. (The Mass psychology of Fascism), (V. R. Cartagno, trans.). N. Y.: Farrar, Straus, and Giroux, 1970. (Originally published, 1933).

Roberts, D. F. and Bachen, C. M. Mass communication effects. In M. Rosenzweig and L. Porter (Eds.), Annual review of psychology (Vol. 32). Palo Alto, Calif: Annual Reviews, 1981.

Snyder, M. L. and Wicklund, R. A. Prior exercise of freedom and reactance. Journal of Experimental Social Psychology, 1976, 12, 120-129.

Sun Tzu. (The Art of War) (S. B. Griffith, trans.). London: Oxford University Press, 1963. (Originally published c. 400 BC).

Tesser, A. Self-generated attitude change. Advances in Experimental Social Psychology, 1978, 11, 289-338.

The stormy summit. The New York Times, November 25, 1981, p. A23.

Toch, H. The social psychology of social movements. N. Y.: Bobbs-Merrill Co., Inc., 1965.

Vaky, V. P. Hemispheric relations: "Everything is part of everything else". Foreign Affairs, 1980, 59, 617-647.

Wood, W., and Eagly, A. H. Stages in the analysis of persuasive messages: The role of causal attributions and message comprehension. Journal of Personality and Social Psychology, 1981, 40, 246-259.

Footnote

The author is indebted to Lt Colonel Norman L. Pfeifer for providing an environment conducive to creativity at Det 1, AFCOS and to Mrs. Sharon E. Ely for furnishing her outstanding administrative expertise.

TEAM TRAINING RESEARCH: INTO THE UNKNOWN?

Jean L. Dyer

US Army Research Institute for the Behavioral and Social Sciences
Field Unit
Fort Benning, Georgia

About 20 to 30 years ago, several major research efforts (see review by Wagner, Hibbits, Rosenblatt & Schulz, 1977) focused on the nature of teamwork and team training within military units. Unfortunately, researchers were unable to adequately follow-up on much of this research, and for a period of time team research was almost non-existent. Nevertheless, much was learned in these early research efforts. This paper focuses on team training recommendations made on the basis of prior research, new data that expand on some of these issues obtained as part of the team research program at the Army Research Institute Field Unit at Ft. Benning, Georgia, and some major questions that still remain unanswered.

Seven recommendations that have been made regarding team training will be examined. These are:

- a. Team training should be preceded by individual training.
- b. Team training should be sequenced in terms of increasing complexity and degrees of teamwork.
- c. Military teams should be trained in conditions which approximate the situations in which they will be expected to perform.
- d. Team training should be conducted periodically.
- e. Team members should receive performance feedback.
- f. Interdependencies among team members should be clarified during team training.
- g. Team training should include training individuals to analyze their own errors, to sense when the team or team members are overloaded, and to adjust their behavior when overloads occur.

Although many of these recommendations seem "obvious", they are not necessarily followed in practice.

"Team training should be preceded by individual training" is a logical recommendation since some degree of individual proficiency is required in most team tasks. In addition, cross-training of skills is not effective until individual expertise is acquired. Experts have stated that the relative emphasis given to each form of training depends on the team task, with tasks that demand little member coordination requiring less team training than those high on this dimension. However, one of the barriers to research in this area is that clear distinctions have not been made between individual and team skills.

Individual and team skills were differentiated in a recent study (Dyer, 1980) of Engineer assault ribbon bridge platoons. Individual skills were defined as activities that could be or were performed independently of other team members; team skills as activities that had to be performed in response to the actions of other team members or that directed the actions of other team members. For example, learning to drive a power boat reflected an individual skill, but operating it in response to the raft commander's directions was defined as a team skill since the boat driver no longer operated as an independent individual. The sequence used in training team and individual skills was observed. In general, individual skills were trained before team skills, more time was devoted to individual than to team skills, and cross-training was limited to individual skills. Interestingly, the method of training the two types of skills varied. Direct verbal instructions were used to train individual skills while team skills were learned through imitation. Although the individual-team training sequence was appropriate for the mission, the training program clearly emphasized individual skills. However, assembly/disassembly of the assault ribbon bridge is definitely a team task; it cannot be completed by one person or several people operating in isolation. Team skills, such as timing the release of bays and boats, interconnecting the bays, and retrieving the bays, are critical to meeting the time standards established by the Army. The failure to address team skills sufficiently may be partially due to the present stress within the Army on individual skills and the lack of analytical tools to guide training developers in the area of team instruction.

Many questions remain unanswered regarding the integration of individual and team skills. Some are: what level of individual proficiency is required before team training begins, should some team skills be acquired before individual skills, how can one determine which individual skills are prerequisite to team skills, and what individual skills can be acquired in a team training situation.

The second recommendation cited above is that "team training should be sequenced in terms of increasing complexity and degrees of teamwork." This relates to the third recommendation that "military teams should be trained in conditions which approximate the situations in which they will be expected to perform," since in most cases the combat environment is more complex and requires more teamwork than training conditions. The most complex form of team training has usually been judged as situations that include emergent (unexpected, new) tasks such as those encountered in combat. A strong advocate of training under combat-like environments is George (1967). His research indicates that if you expect a team to perform under stress, members must train under stress. If you expect a team to work as a coordinated unit, members must be trained under conditions that force teamwork. The success of training programs such as REALTRAIN also testifies to the importance of realistic training settings (Banks, Hardy, Scott, Kress, & Word, 1977).

Introduction of the Multiple Integrated Laser Engagement System (MILES) into the Army community has enhanced training realism. MILES uses laser transmitters on weapons and laser detectors on targets to determine casualties in force-on-force exercises. Observation of MILES exercises indicates that soldiers do behave differently than in exercises without MILES. For example, the realization that casualties

may occur affects the way Infantry squads move across the terrain and highlights the importance of cross-training and of communicating critical coordination information to the entire squad. In one exercise, the radio operator was killed, but no one else knew the radio signals and the unit was unable to call in indirect fire. In another exercise, the radio operator was again killed, but it took other platoon members a considerable amount of time to read the map and call in indirect fire. In the first case, the squad's chances of succeeding in the mission were greatly reduced; in the second case, valuable time was lost.

However, even a realistic training situation will lose its effectiveness when not implemented appropriately into small-unit training. Obviously, the rate at which the complexity of team training should be increased is a critical research question. Exposing teams to complex situations at either too fast or too slow a rate may be ineffective and/or waste training time.

Repetition of training exercises and feedback during training are critical to training success, even in realistic training settings. In fact, two major training recommendations that have been made and are strongly supported by individual learning theory and research are that "team training should be conducted periodically," and "team members should receive performance feedback." One of the complicating factors regarding feedback is that it can be given at both the individual and the team level. When team feedback is given, individual team members may be reinforced for inappropriate behavior. Briggs and Johnston (1967) have an excellent review of training guidelines in this area, although more research is needed.

The importance of repetition cannot be denied, but too often team exercises are conducted only once. Such small amounts of training do not even insure that team skills are initially acquired. A clear illustration of the positive effect of repetition occurred in a recent study involving a movement to contact mission pitting an Infantry squad against an opposing force of three individuals (Dyer, Bennett, & Sulzen, 1981). All individuals used MILES equipment. Each squad conducted the mission once. However, the opposing force faced 27 squads, 9 per week over a period of three weeks. Two opposing forces participated in the study. With each additional week, the opposing force became more skilled. This was true for each opposing force. During the first week, the opposing force won 39% of the exercises; during the second week, 44% of the exercises; and during the last week, 83% of the exercises.

The last two team training recommendations cited above focus specifically on requirements unique to team training. These recommendations are "interdependencies among team members should be clarified during team training," and "team training should include training individuals to analyze their own errors, to sense when the team or team members are overloaded, and to adjust their behavior when overloads occur." Some research in command and control systems has indicated that the best way to enhance mission performance is to reduce the extent to which coordination among team members is required (Briggs & Johnston, 1967, Roby & Lanzetta, 1967). However, coordination demands cannot be eliminated from all military missions.

A barrier to team research, and ultimately to the development of team training programs, is that "coordination" is poorly defined. One of the major efforts of the team research program at ARI-Ft. Benning is to clarify what is meant by coordination at both the function and the task levels. Initial work at the function level (Dyer et al., 1981; Shiflett, Price & Schemmer, 1981) indicates that it is possible to identify and measure different aspects of coordination, such as orienting the team to the mission, distributing/redistributing equipment, assigning/reassigning tasks, pacing and sequencing team activities, and adjusting team member responses to the actions of other members. Observational studies have indicated, however, that team members often fail to coordinate. For example, Infantry rifle squad operation orders (Dyer, et al., 1981) gave very little attention to planning for team member coordination during the mission, and such coordination was not covered by squad standing operating procedures (SOP). Few contingency plans were given. Critical information about the nature of the enemy was omitted. Briefing modes were used which decreased the likelihood that all squad members would get the same information regarding the mission. During execution of a Military Operations in Urban Terrain (MOUT) exercise (Dyer, 1981), coordination among team members was minimal. For example, rifle platoon members failed to communicate between the assaulting squad and the fire support squad, and members failed to consider the fact that their bullets could hit other members on different floors of the building that was being cleared.

Many research questions need to be addressed regarding teamwork and coordination. Very little is known about how coordination patterns change as a team becomes more skilled, what types of coordination are the most difficult to acquire, what training techniques are effective in teaching different forms of coordination, and how much repetition is required for the retention of team skills.

In answer to the question raised in the title of this paper, neither team researchers nor training personnel need start in a vacuum. Implementation of the training recommendations cited at the first of the paper, followed by systematic research on related but unanswered questions and application of the research results, could have a major impact upon the effectiveness of small-unit training within the military.

References

1. Banks, J. H., Hardy, G. D., Scott, T. D., Kress, G., & Word, L. E. REALTRAIN validation for rifle squads: Mission accomplishment (ARI Research Report 1192). Alexandria, Va.: US Army Research Institute for the Behavioral and Social sciences, October 1977. (DTIC No. AD 043 515)
2. Briggs, G. E., & Johnston, W. A. Team training (Technical Report, NAVTRADEVCEEN 1327-4). Columbus, Ohio: Ohio State University, Human Performance Center, June 1967. (DTIC No. AD 660 019)
3. Dyer, J. L. The initial training of individual and team skills: An exploratory investigation of Engineer bridge specialists (ARI Working Paper FBNG FU 80-1). Fort Benning, Ga.: Army Research Institute for the Behavioral and Social Sciences, Fort Benning Field Unit, October 1980.
4. Dyer, J. L. Prediction of Infantry squad errors during training: Pilot investigation (ARI Technical Report). Alexandria, Va.: US Army Research Institute for the Behavioral and Social Sciences, July 1981 (in press).
5. Dyer, J. L., Bennett, R., & Sulzen, R. Infantry rifle squad operation orders: Their characteristics and role in mission success. Unpublished manuscript, 1981. (Available from ARI Field Unit, Ft. Benning, Ga.)
6. George, C. E. Training for coordination within rifle squads. In T. O. Jacobs, J. S. Ward, T. R. Powers, C. E. George, & H. H. McFann (Eds.) Individual and small-unit training for combat operations (HumRRO Professional Paper 21-67). Alexandria, Va.: Human Resources Research Office, May 1967. (DTIC No. AD 653 845)
7. Roby, T. B., & Lanzetta, J. T. Conflicting principles in man-machine system design. Journal of Applied Psychology, 1957, 41, 170-178.
8. Shifflett, S., Price, S. J., & Schemmer, F. M. The identification and measurement of team functions. Paper presented at the meeting of the American Psychological Association, Los Angeles, August 1981.
9. Wagner, H., Hibbits, N., Rosenblatt, R. D., & Schulz, R. Team training and evaluation strategies: State-of-the-art (HumRRO TR-77-1). Alexandria, Va.: Human Resources Research Organization, February 1977. (DTIC No. AD A038 505)

A TAXONOMIC MODEL FOR NAVY TEAMS

LARRY NADLER
LITTON MELLONICS

ABSTRACT

This paper documents research conducted by Litton Mellonics for the Office of Naval Research. The primary purpose of this research entailed the development of a Navy team taxonomy. This taxonomy was designed to refine the notion of Navy teams, establish a framework for systematic military team research, aid scientists in selecting teams for research, and clarify the applicability of team and small group research to military performance issues. The nature of Navy teams and taxonomic structures was explored and a systems model was employed to organize several team-related variables. From the literature review of relevant team research, the taxonomic model was generated. Measures of the taxonomic dimensions were developed and applied to a sample of 245 surface Navy teams.

A TAXONOMIC MODEL FOR NAVY TEAMS

INTRODUCTION

After reviewing the state of military training of teams, the Defense Science Board (1976) asserted the importance of developing a naval team taxonomy. This taxonomy is needed to refine the notion of naval teams, to establish a framework for systematic military team research, to aid scientists in selecting teams for research, and to clarify the applicability of team and small group research to military team performance issues. In addition, this taxonomy represents an important step toward developing improved measures of the dimensions of naval team performance.

There were two major objectives of this research effort. First, it was necessary to design a taxonomy, or classification system, which captures significant variations along pertinent dimensions of naval teams. To accomplish this task, a workable operational definition of "team" and an examination of the relevant theoretical and experimental literature to identify salient team dimensions are required. Second, this taxonomic model was applied to surface Navy teams to generate a catalog which facilitates team selection for diverse research purposes. The accomplishment of these task objectives entailed an integration of the team performance literature with specific technical attributes of surface Navy teams.

THE DELINEATION OF NAVY TEAMS

An important, yet difficult, feature of this research effort was the delineation of naval teams. Meister (1976) captured this problem in the following manner: if the focus is too narrow, the definition of teams would exclude supervisory personnel, whereas if the scope is too broad, the team description would include outside influences. There are many other reasons why the identification of Navy teams is an arduous process, including overlapping team memberships, borderline members, the instability of team composition and structure, and the existence of "distinct subsidiary teams" (i.e., subteams) within identified teams. Clearly, an important product of this research is the specification of appropriate levels of analysis to catalog and examine naval teams.

As naval teams can be discerned at various levels, a hierarchical delineation comprised of successively narrower units is most appropriate in depicting these teams. In other terms, naval teams were described at multiple levels, permitting researchers to select the breakdown of teams most relevant to their particular investigations. At the broadest level, an entire ship can be construed as a team. Further, the personnel working in a specific "functional area" can be viewed as comprising a team. The seven functional areas identified are seamanship, combat information, weapons, communications, engineering, damage control, and support. These functional groups are comprised of smaller teams, which constitute the third level of analysis. For instance, within the seamanship area, one can identify conning, navigation, anchor handling, boat handling, boat operating, and cargo handling teams.

In delineating Navy teams, an examination of various definitions of teams in the literature was performed to identify critical elements. The major dimensions identified by these definitions are number of members, synergistic relationships, task interdependence, cooperation/coordination requirements, team structure, and normative prescriptions. The primary data source for delineating a representative set of surface Navy teams involved the corresponding selection of sixteen Ship Manpower Documents. Having considered the theoretical and practical aspects of defining and delineating Navy teams, the next research task involved the identification of salient team dimensions and the generation of an appropriate taxonomic model.

THE TAXONOMIC MODEL

While various theoretical models have been proposed to describe the relationship between team characteristics and performance, a systems approach seems to capture best the dynamic nature of teams and the variables which impinge upon team productivity. Typically, systems models organize team variables into input, process, and output categories. Input variables describe the initial state of a task-oriented group and they include organizational/environmental/situational, individual member, and team-specific factors. Team interaction process, according to Collins (1977), entails "all observable interpersonal behavior that occurs between two arbitrary points in time ..." (p. 3-39) and mediates between input and output. Output variables include task performance and interactive components (e.g., member satisfaction). Comprehension of a team's task-related productivity hinges upon assessing the impact of input and process variables on team output.

A major advantage of systems models of team performance involves their ability to synthesize many team variables into logical categories and describe the relationships among the categories. Thus, such models provide the grouping of the essential characteristics of teams, a grouping that is required for the design of a team taxonomy. As such, this project attempted to identify key team characteristics, with the major criterion for inclusion in the model being a demonstrable link to team performance. The major taxonomic elements, based upon an extensive literature review, were team tasks, team members, team structure, leadership, and team communication processes.

Team Tasks

An important dimension of the productivity of naval teams entails characteristics of the tasks which they perform. One important dimension of team tasks is task difficulty, which Shaw (1976) defined as the "amount of effort required for task completion" (p. 312). Daniels and Alden (1975), employing a systems perspective, suggested that task difficulty manifests itself differently during input (e.g., stimulus uncertainty), process (e.g., cognitive information-processing demands), and output (e.g., response complexity) activities. In a naval setting, task difficulty can involve the operation of highly automated equipment, the evaluation of complex tactical data, the utilization of highly technical symbolic coding in sending/receiving messages, etc. Another key task characteristic is task emergence. Boguslaw and Porter (1962) distinguished between established and emergent situations by referring to the extent to which action-relevant environmental conditions and states are specifiable and predictable. When unanticipated, emergent situations arise, the coordination demands placed on the team increase and may influence performance. Among the system conditions most relevant to naval team functioning are stimulus uncertainty/complexity, environmental conditions (e.g., weather), equipment failure, and battle casualties. A third task dimension is task type. Steiner (1966, 1972) distinguished between disjunctive, conjunctive, complementary, compensatory, additive, and discretionary task situations, noting that these task types differ in the nature and extent to which team members are permitted or required to combine resources for mission completion. The final two task features included in the team taxonomy are task content and machine interface. The former variable refers to the team's task performance criterion, while the latter dimension entails the degree of automation of the equipment which the team utilizes for mission accomplishment.

Team Member Characteristics

A second category of input variables focuses upon team member characteristics. Two important team member attributes are proficiency and experience levels. According to Meister (1976), the proficiency of team members is intimately linked to effective team performance, accounting for as much as fifty percent of the variance in team productivity. In

addition to task-related capabilities, other factors, such as personality attributes and motivational levels, are capable of affecting team performance. Another important dimension of member characteristics, in an accumulative sense, is team size. As team size increases, "the range of abilities, knowledges, and skills that are available to the group increases... as well as the sheer number of 'hands' that are available for acquiring and processing information" (Shaw, 1976, p. 155). However, Hackman and Vidmar (1970) noted that increases in team size are frequently accompanied by greater communication difficulties, less intermember cooperation, and members' feelings of possessing less influence. The specific effects which team size has upon productivity are related to several other factors, including the nature of the task, group structure, and member proficiencies. A final team member attribute entails member criticality. While it is assumed that all individuals are needed for task completion, it is acknowledged that criticality asymmetry exists and may influence team performance, especially when battle casualties or manpower shortages occur.

Team Structure

According to Davis (1969), team structure is a "picture of the interpersonal processes among the positions taken at a particular point in time" (p. 88). In applying the group structure literature to naval teams, it is necessary to consider the nature of these teams. As Davis indicated, team structure in military organizations is formal and imposed. Further, Glanzer (1962) observed that military teams are typically highly centralized, although not always in a communicative or interactive sense. In this context, the taxonomic model focuses upon the serial versus parallel structure; whether interactions are face-to-face, by audio, and/or by video means; the nature of the decision-making structure; and the informal structure adopted by Navy teams.

Leadership

The final input category involves leadership. At the simplest level, leader identification is a function of formal naval designation. Generally, Navy documents emphasize task-oriented, socio-emotional, and role model functions served by naval leaders. Other leadership variables include leadership style, such as Lewin, Lippitt, and White's (1939) distinction between democratic, autocratic, and laissez-faire leadership styles, and leader-member relations, a specific feature of Fiedler's (1964, 1967) contingency approach to effective leadership behavior.

Team Communication Processes

While various approaches to classifying interaction behaviors have been adopted, two categorization schemes merit attention. Bales (1950) employed Interaction Process Analysis to classify communication behaviors into twelve categories which reflect task-oriented and social functions. More recently, Nieva, Fleishman, and Rieck (1978) developed a communication process typology consisting of four areas of communication process variables. These include orientation (i.e., information distribution); organization (i.e., coordination processes); adaptation (i.e., cooperation processes); and motivation (i.e., the generation and enforcement of team norms). In addition, team cohesiveness, defined by Shaw as "the resultant of all those forces acting upon group members to remain in or leave the group" (p. 446), is clearly a function of task-related and social communication outcomes.

THE CLASSIFICATION PROCESS

The development and application of the taxonomic measures can be described. These measures should permit critical variations between different teams and within the same team type across different ships to emerge. Further, the taxonomic measures should thoroughly capture aspects of Navy teams which are demonstrably related to the performance of these teams. A significant constraint on the achievement of these objectives, though, involved the inaccessibility of specific teams in operational settings. Thus, a distinction between

exogenous and endogenous team dimensions was warranted. Exogenous dimensions characterize teams without regard for the individual attributes and behaviors of the actual team members, while endogenous dimensions entail team-specific features. Although these latter factors are important descriptors of Navy teams, their measurement was severely limited for various reasons. Specifically, they represent unique variance within teams, thus requiring detailed observation of ongoing team communications and performance. Also, this individual team variance would limit the generalizability of results because extrapolations from observed teams to other similar teams involve considerable measurement error.

Data Sources and Sampling

A variety of data sources was used to obtain the ratings for the measures of the taxonomy. Documentary sources included Ship Manpower Documents, training course materials, ship and system design studies, research studies, tactical doctrine publications, and Navy regulations. Other sources of information included subject matter experts at Navy team training sites and direct observations of teams in training situations.

The ships in the sample were chosen to provide a diversity of sizes, capabilities and systems within the sample. Also, we sought to select ships which are representative of the population of ships outside the sample. Overall, 245 teams from 16 ships were catalogued within this research effort.

The Taxonomic Measures

At this point, the taxonomic measures can be introduced and an illustration of their application provided. As indicated earlier, the team is delineated at the ship, functional area, and individual levels. Team members are identified and proficiency/experience levels can be inferred from the rate/designators and NEC/NOBC's. Additionally, criticality ratings and team size (by readiness condition) are identified. Task characteristics, team structure variables, the nature of team leadership, and the importance of team interaction processes for task completion are also assessed.

DISCUSSION

In providing a common focus regarding Navy teams, dimensions, and corresponding measures, this research offers an organizing framework for construing prior studies and designing subsequent research endeavors. Retrospectively, the taxonomic model allows any research effort to be understood in terms of the teams examined and the variables which were manipulated, controlled, or unaccounted for. Prospectively, future research can be predicated upon enhanced understanding of the critical features within a particular research environment.

A unique aspect of this research is that the classification system represents an integration of the massive team performance literature with the specific technical aspects of Navy teams. This integrative effort serves the important purpose of assisting Navy researchers in selecting teams based upon functional similarities/differences across various taxonomic dimensions. Further, the application of the taxonomic measures to Navy teams generated an understanding of those dimensions which require more complex procedures to measure meaningfully. Specifically, the identification of endogenous team dimensions indicated that field observations are most appropriate for their measurement, while exogenous features can be assessed using manning documents.

The classification system possesses other forms of utility for team investigators. For instance, increased understanding of the ways in which Navy teams differ can facilitate the design of team-specific training programs. To the extent that current team training does not consider important differences between Navy teams, beneficial modifications could be made. For example, a Navy team which relies heavily on communication and coordination should receive training in these interactive skills. Also, an analysis of member criticality ratings could

lead to the differential imparting of cross-training to specific team members. Further, in the context of manpower shortages, less capable recruits, and more complex machine systems, the taxonomic model could be employed to suggest optimal utilization of team resources. In our opinion, the major utility of this research effort resides in the ability of other team investigators to make beneficial refinements and identify areas of application oriented toward improving the performance of Navy teams.

References

- Bales, R. F. Interaction process analysis: A method for the study of small groups. Cambridge, MA: Addison-Wesley, 1950.
- Boguslaw, R. R. and Porter, E. H. Team functions and training. In Gagne, R. M. (ed.), Psychological principles in system development. New York: Holt, Rinehart, and Winston, 1962.
- Collins, J. J. A study of the potential contributions of small group behavior research to team training technology development. (Final Report ONR-00014-76-C1076, NR-179-34). Alexandria, VA: Essex, August 1977.
- Daniels, R. W. and Alden, D. G. The feasibility of generalized acoustic sensor operator training. Minneapolis, MN: Honeywell, Inc., Systems and Research Center, May 1975.
- Davis, J. A. Group performance. Reading, MA: Addison-Wesley, 1969.
- Defense Science Board. Crew/group/unit training. In Report of the Task Force on Training Technology. Washington, D. C.: Office of the Director of Defense Research and Engineering, Department of Defense, February 1976.
- Fiedler, F. E. A contingency model of leadership effectiveness. In L. Berkowitz (ed.), Advances in experimental social psychology, Vol. 1. New York: Academic Press, 1964.
- Fiedler, F. E. A theory of leadership effectiveness. New York: McGraw-Hill, 1967.
- Glanzer, M. Experimental study of team training and team functioning. In Glaser (ed.), Training research and education. Pittsburgh: University of Pittsburgh Press, 1962.
- Hackman, J. R. and Vidmar, N. Effects of size and task type on group performance and member reactions. Sociometry, 1970, 33, 37-54.
- Lewin, K. Lippitt, R. and White, R. K. Patterns of aggressive behavior in experimentally created "social climates". Journal of Social Psychology, 1939, 10, 271-199.
- Meister, D. Behavioral foundations of system development. New York: John Wiley, 1976.
- Nieva, V. F., Fleishman, E. A. and Rieck, A. Team dimensions: Their identity, their measurement and their relationships. Advanced Research Resources Organization, Princeton, NJ: Response Analysis Corp., November 1978.
- Shaw, M. E. Group dynamics. New York: McGraw-Hill, 1976.
- Steiner, I. D. Models for inferring relationships between group size and potential group productivity. Behavioral Science, 1966, 11, 273-283.
- Steiner, I. D. Group process and productivity. New York: Academic Press, 1972.

A METHOD FOR DESCRIBING TEAMS AND TEAM BEHAVIOR¹

Rohn J. Hritz
Thomas J. Roth
Donald W. McGill

Applied Science Associates, Inc.
Valencia, Pennsylvania

and

Jean L. Dyer

Army Research Institute
Fort Benning, Georgia

A descriptive model of team behavior and performance has been developed by Applied Science Associates, Inc. under contract to the Army Research Institute, Fort Benning, Georgia. The descriptive model identifies a set of dimensions useful in describing teams and team behavior. A method was developed for systematically applying the descriptive model concepts to a given team in order to generate team structure and behavior descriptions. The method can be used to generate descriptions for any type or kind of team performing any team task. The purpose of this presentation is to present the significant features of this method.

Development of the Method

The method was developed directly from the descriptive model of team behavior. Thus, it is important to briefly discuss how the model was developed. The descriptive model was developed and verified by observing six selected Army teams performing selected missions or tasks. Four waves of observations were conducted. After each wave, descriptive concepts were identified and/or further developed. The first three waves of observations involved five different team types (Infantry rifle squads, Improved TOW Vehicle (ITV) squads, mortar squads, mortar fire direction centers, and assault ribbon bridge platoons). Four of the teams were observed in each of the three waves of observation. The assault ribbon bridge platoon was observed only during the second wave of observations. Each of the teams was observed performing approximately five different missions.

The fourth wave of observations was reserved for verifying the identified concepts. This wave involved two team types; one which was observed in all three previous waves (the Infantry rifle squad), and one team type not previously observed (Field Artillery fire direction center). The previously observed team type was included in the verification observation to provide a check on the accuracy of the identified concepts. The new team type was included to

¹ This research was supported by a contract from the Army Research Institute Field Unit, Fort Benning, Georgia (Contract No. MDA390-8a-C-0198). The opinions expressed in this presentation are solely those of the authors and not necessarily those of the Army Research Institute or the Government.

provide a check on the completeness of the model. The model would be judged incomplete or inaccurate if the identified concepts could not be applied to either team or if, for the new team, the model concepts could not describe phenomena that were observed.

Once the model and its concepts were verified, the descriptive method was developed. The method consists of step-by-step procedures or instructions for describing any team performing any of its missions. The descriptive method was verified by applying it to a new team type (not previously observed in model development). During the verification, two independent observers each generated four independent team descriptions; one for each selected mission performed by the team. The resulting descriptions were then compared and the discrepancies noted. The discrepancies were traced back to the step-by-step instructions (i.e., the method). Revisions were then made in the method to minimize any misinterpretation of the instructions or procedures. In addition, the resulting descriptions were verified for accuracy by subject matter experts familiar with the team type and the selected missions.

Descriptive Method

The descriptive method can be applied using data gathered from any or all of the following sources: (a) documentation about the team (e.g., field manuals), (b) observations of the team performing the missions of interest, and (c) interviews with subject matter experts (SMEs).

Information about the team and its behaviors are recorded on a set of forms. Each form is accompanied by a set of instructions. The method also contains a series of decision aids. The decision aids provide guidance in determining how specific concepts of the descriptive model are to be applied. For example, the descriptive model introduces the concept of nested teams. This concept is not appropriate in all situations, so a decision aid was developed to assist the user of the descriptive method in determining where the concept is appropriate.

The descriptive method requires two kinds of data to be collected; data describing team structure and scope, and data describing the behaviors of the team as a specific mission is performed. The data collected consist of the following:

1. A list of all the missions (team tasks) performed by the team.
2. A description of the positions that comprise the nominal structure of the team. The nominal team structure is that dictated by Tables of Organization and Equipment (TO&E). Nominal team structure reflects the organization of the team when it is neither overstrength nor understrength and has all of its required equipment. Nominal team structure is an out-of-mission-context description.
3. A list of equipment used by each team member.
4. An organizational chart indicating any hierarchical relationships among team members.

The above data are designed to provide an overview of the basic resources of a team.

Once the overview data have been gathered and recorded on the appropriate forms, each mission of interest performed by the team is then described. The following data are collected for each mission:

1. The name or title of the mission.
2. A brief narrative description of the mission. This narrative includes a statement concerning the goal of the mission, plus a general description of team activities.
3. A topographic map of the mission. The map accompanies the narrative, and indicates the following: troop movement, opposing force movement, terrain characteristics, location of cover and concealment, terrain elevations, location of vehicles, etc.
4. A description of the conditions surrounding the performance of the mission, e.g., the use of special equipment, the use of non-organic mission support (such as supporting artillery fire), and the environmental factors influencing the attainment of the mission (such as time of day, climate, etc.)
5. A description of the actual team structure used to accomplish the mission. The actual team structure may be different from the nominal team structure, since the team may be overstrength or understrength. The actual team structure is described by listing the team members and specifying the equipment used by each member. In addition, the actual team structure is described with an organizational chart.
6. A brief description of the function performed by each team member in terms of the team task, i.e., the contribution made by each member to the successful completion of the team task.
7. A description of the team behaviors required in the performance of the mission. This includes a description of the dependencies that occur among team members. A dependency is defined as the exchange or transfer of information, products, or services among team members.

All the dependencies (team behaviors) are recorded on a mission description form. This form contains one column for each team member, plus a column for comments. Information about each dependency is entered in the column corresponding to the team member who initiated the dependency and in the column of the team member or members who were recipients of the dependency. In the initiator column, the following information is entered:

1. The activity the initiator was engaged in when the dependency was precipitated, i.e., where in his or her individual tasks did the dependency occur.

2. A symbol for the dependency pattern. A dependency pattern consists of a diagram indicating the number and identity of initiators and recipients. A decision aid has been developed to help the analyst determine the most appropriate dependency pattern.
3. A description of the dependency element, i.e., a statement describing what occurred between the initiator(s) and the recipient(s).
4. A description of the dependency type, i.e., a code indicating if information, a product, or a service was transferred. The descriptive model contains three major classes of dependency type. Each major class is divided into subclasses. A decision aid has been developed to assist the describer in selecting the most appropriate dependency type.
5. A description of the purpose of the dependencies, i.e., a code indicating the purpose of the dependency. Approximately nine generic purposes were identified during initial observations. The analyst selects one of these nine purposes.
6. A consequence rating. The rating ranges from one to four indicating the effect of a failure to execute the dependency upon completion of the team mission. A procedure for assigning this rating is provided as part of the method.

In the recipient's column the following information is entered about a dependency:

1. The description of the activity the recipient(s) was (were) performing when the dependency element was received. This information describes where the dependency occurs in the recipient's individual tasks.
2. A brief description of the status of the dependency element, i.e., a statement describing the state of the element when it is received.

The comments column is reserved for recording any additional information about the dependency.

Each dependency is described in a separate row on the form. Thus the user can obtain an overview of the entire team task by scanning the recording form.

In summary, the method yields a team description for a particular mission that includes both the nominal and actual structure of the team, the equipment the team should have for the mission as well as the equipment actually available, narrative and pictorial descriptions of the mission, and the dependencies among the team members as they perform the mission including the sequence in which the dependencies occurred, the initiator(s) and recipient(s) of the dependency, the dependency type, its purpose, and the consequences of failure to complete each dependency.

Uses of the Descriptions

The descriptions produced from the application of the method have many potential uses. Of interest at the present time is using the method to identify team training requirements and to develop team training programs. From the descriptions, it will be possible to identify training objectives and to note the training deficiencies of specific teams.

AN EMPIRICALLY DERIVED DESCRIPTIVE MODEL
OF TEAM BEHAVIOR: THE CONCEPT OF DEPENDENCY¹

Rohn J. Hritz
J. Thomas Roth
Donald W. McGill

Applied Science Associates, Inc.
Valencia, Pennsylvania

and
Jean L. Dyer

Army Research Institute
Fort Benning, Georgia

The training challenge of the 1980s is the advancement of team training technology and the improvement of team performance. In the past decade, the military has concentrated its efforts on the development and delivery of individual training. However, many combat missions are performed by a collection of highly trained individuals working together as a team. Team members, often with diverse skills and knowledge, must coordinate their efforts and communicate and interact with each other in order to successfully accomplish their team task. In this situation, providing individual training alone may not be enough to guarantee a satisfactory degree of team proficiency.

Unfortunately, little of practical value is known about team behavior and team performance. Most prior research has emphasized the study of small groups (not teams) formed specifically for experimental purposes, to work on experimentally contrived intellectual tasks. The kinds of teams and team tasks reported in the literature are not representative of the kinds of teams and team tasks encountered in the military. Furthermore, prior research has produced contradictory results, indicating that no consistent theory of team behavior exists.

The absence of a theory of team behavior is, in part, due to the lack of a conceptual framework for describing team behavior. One cannot expect to explain and predict team behavior without first possessing the ability to describe team behavior. This presentation discusses a model for describing team behavior. In particular, a unified concept of team member interaction and interdependency is presented. This concept is known as dependency.

¹ This research was supported by a contract from the Army Research Institute Field Unit, Fort Benning, Georgia (Contract No. MDA390-81-C-0198). The opinions expressed in this presentation are solely those of the authors and not necessarily those of the Army Research Institute or of the Government.

Development of the Model

The descriptive model was developed by observing selected Army teams performing selected team tasks or missions. None of the teams or missions observed was contrived for the purposes of the study. All the teams were studied in field situations rather than laboratory settings.

The observations occurred in four waves or phases. After each phase of observation, concepts concerning team behavior were identified and further refined. Thus, the descriptive model was developed and validated in an iterative fashion.

Team types varying along several significant dimensions were selected for study. The team types varied in size (from four to twenty-two members), in their reliance upon and use of equipment, and in the type of missions performed. The following team types were observed: Infantry rifle squads, Improved TOW Vehicle (ITV) squads, mortar squads, mortar fire direction centers, and assault ribbon bridge platoons. Each team type, except the assault ribbon bridge platoon, was observed in each wave of observation. The assault ribbon bridge platoon was only available for observation during the second wave of observation. Each team was observed performing approximately five different missions. Each wave of observation involved a different specific team of each team type, i.e., observations were not made on the same team, just the same team type.

The model was validated by observing a team type previously observed (Infantry rifle squad) as well as a team type not observed during the development of the model (Field Artillery fire direction center). If previously identified concepts were inappropriate to describe either team type, then the accuracy of the model would be suspect. If model concepts were not sufficient to describe the "new" team type, then the completeness of the model would be questioned. Although the model was refined after the validation exercise, no new concepts were added.

The descriptive model contains several important concepts. One of the most useful concepts is that of a dependency. In fact, this presentation will not deal with the entire model, but will only describe the dependency concept. This concept is extremely useful because it provides guidance in differentiating between individual behavior and team behavior.

The Dependency Concept

While observing various team types, it became clear that one of the phenomena that made a collection of soldiers into a team was their dependence upon each other. In order to successfully achieve individual missions, each team member depends upon other members for information, products, and services. Each team member is, of course, responsible for completing individual tasks. The completed tasks contribute to the success of the team mission. Individual tasks, however, often require input in order to be completed. In addition, each individual task generates at least one output. A dependency is said to exist when inputs and outputs are exchanged or transferred among team members during a mission.

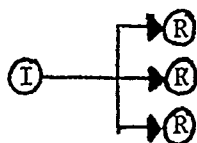
There are two classes of dependencies, internal and external dependencies. An internal dependency exists when a member of the team requires an input from another team member in order to complete his individual task or when he generates an output required by another team member. An external dependency exists when an input is received from or an output is used by someone outside the team's organization structure. Internal dependencies (dependencies among members of the team) are considered one factor or dimension of team behavior. They are the relational links among the team members. If an internal dependency is not properly executed, then the team cannot successfully and efficiently complete its team mission or task.

Thus, the dependency concept not only defines team behavior, it helps to differentiate individual behavior from team behavior and, to some extent, it helps to define what constitutes a team. If a collection of individuals has no internal dependencies, then it is questionable whether, in fact, a team exists, since members do not depend upon the outputs of each other's activities to accomplish a team mission.

Description of Dependencies

The descriptive model of team behavior specifies how dependencies should be described. Every dependency within the team contains three elements. These are: an initiator, a recipient, and the dependency element. A dependency element is defined as the information, product, or service that is exchanged or transferred among the participants in the dependency. In the simplest case, there is one initiator, one recipient, and one dependency element. For example, when a squad leader gives an order to one rifleman, there is one initiator (the squad leader), one recipient (the rifleman), and one dependency element (the order). Observations conducted during the development of the model allowed the identification of a number of kinds of dependency patterns that involved more than one dependency element or multiple initiators or recipients.

Each of the dependency patterns has been precisely defined, based upon what was observed in the field. For example, a "fan-out" pattern is characterized by a single initiator, multiple recipients, and one or more related dependency elements. A fan-out pattern can be diagrammed as:



A fan-out pattern describes the dependencies (for example) between the team leader (initiator) and the team members (recipients) when an operation order is given (the content of the operation order comprises the dependency element).

Not only can a dependency be described by the initiator, recipient, dependency element and pattern, but it can also be described by its type. Field observations revealed several types of dependencies: communicative dependencies, product dependencies (both physical product and procedural), and virtual dependencies. Each of these types has been precisely defined.

A communicative dependency exists when the dependency element transferred or exchanged consists solely of information. There are two types of communicative dependencies: verbal and non-verbal. Verbal dependencies include speech and writing, either non-mediated or mediated via equipment (e.g., teletype, computer message, etc.). Non-verbal communicative dependencies consist of signs and signals (e.g., hand signals).

Product dependencies are those where the dependency element consists of a physical product or a logical product. For example, in a mortar squad, the ammunition bearer prepares a round and passes the round to the assistant gunner who inserts it into the mortar tube. This is an internal dependency with one initiator and one recipient. The dependency element is the transfer of a physical product, the prepared round. A logical product or procedural dependency exists when the dependency element is the completion of a procedure or set of activities that must be completed by the initiator before the recipient can begin or continue performing his individual task. A procedural or logical product dependency involves the transfer of a service and not a physical product.

Virtual dependencies are those where the dependency element is implied and not directly observable, i.e., where the product or service transferred is implied and not made explicit by either the initiator or recipient. Perhaps the best way to explain virtual dependencies is by example. Infantry squad members, on a movement to contact mission, are all responsible for maintaining separation between individuals and for surveillance of the area in their section, i. e., each member depends upon the others to provide security for themselves and the other team members. Thus, a service is transferred among all the members, but the service is never explicit, nor is the service required for the other team members to continue their assigned individual tasks. It is this characteristic which differentiates logical product dependencies from virtual dependencies. Virtual dependencies also involve the sharing of a set of tasks in which all participants perform part of the tasks without being explicitly assigned a specific subset of tasks. Virtual dependencies have been observed during the erection of camouflage and during the connection of ribbon bridge units. These involve virtual dependencies because all participants know what needs to be accomplished, and the order in which they are to be performed without being specifically instructed, i.e., in the process of completing the team task, team members seem to survey what has been done and automatically perform what has not already been performed by some other member.

Dependencies can also be described by purpose. Observations suggest that dependencies occur for several reasons. An initial set of dependency purposes derived from observation is: (1) to transfer orders, instructions, or directions, (2) to transfer products, (3) to provide mission status data, (4) to provide status concerning the completion of an individual task, (5) to indicate personnel status, (6) to provide status of supplies and equipment, (7) to provide feedback or corrective actions, (8) to provide information about the mission environment, and (9) to provide information about the enemy. Virtual dependencies can have either a context purpose or a purpose unique to the completion of the task.

Dependencies can also be described by their consequences, i.e., the consequences to a mission from a failure to execute a dependency. Currently the consequence of a dependency failure is made on a four-point rating scale. The rating is determined by analyzing the context of the dependency.

Summary

Internal dependencies are a critical dimension of team behavior. Internal dependencies are defined as the transfer of information, products, or services among team members. To some extent, internal dependencies define the existence of a team, in that a collection of individuals which has no internal dependencies is probably not a team. A dependency can be described by specifying the initiator, the recipient, the dependency element, the type of dependency, the purpose of the dependency, and the consequences involved in the failure to execute the dependency. The concept of dependency, as well as the classes of dependency types and dependency purposes, were derived by observing actual Army teams performing actual missions.

STRESS MANAGEMENT: A MODEL FOR OPTIMIZING HUMAN RESOURCE EFFECTIVENESS

*T. Roger Manley
Florida Institute of Technology*

ABSTRACT

The author recently conducted nine stress management workshops on a major DoD reservation. The courses have been very favorably received and command level attention has been given to data collected in the workshops. This paper describes a general four step model used in workshop design, stress measurement techniques and results, and offers suggestions for behavioral scientists interested in conducting similar workshops.

Introduction

In the Spring of 1981 the civilian personnel office on a major DoD installation in the Southeastern United States advertised a two day professional development course on stress management. Response to the notice of the course offering was overwhelming. The workshop was oversubscribed in a matter of two days, requiring the scheduling of further offerings. In all, a total of nine were conducted for 231 participants.

Much has been written in recent times about job related stress. The subject was a major subtheme in last year's air controllers' strike, and it often seems that one can hardly get through the week without encountering at least one report of some research effort which has explored yet another aspect of job related stress and its association with a widening variety of somatic ailments. One team of researchers recently went so far as to suggest, presumably half in jest, that laws requiring hazard warning on product labels be applied to certain careers (Kreitner, Wood and Friedman, 1981). More recently stress management has been discussed in the literature as an OD intervention (Adams, 1978) and as an area of emphasis for more effective human resource management (Ivancevich and Matteson, 1980).

Workshop Design

In designing the workshop we first recognized the ubiquity of stress. Selye's (1974) definition of stress ("... the nonspecific response of the body to any demands made upon it") served as a focal consideration in determining what should be included. We agreed completely with Dr. Jane Templeton's observation that stress is "the Purple Heart of an eventful life ... more specifically, your body's physical, mental, and chemical reactions to all the things around you" (Ivancevich and Matteson, 1980). Based upon the foregoing, a four phase stress management approach (Figure 1) was adopted: (1) understanding the nature of stress (i.e., what is stress; where does it come from; and what effect does it have on your body); (2) identifying symptoms of stress in ourselves and others; (3) learning about and experimenting with various stress management techniques; and (4) developing a personal stress management plan.

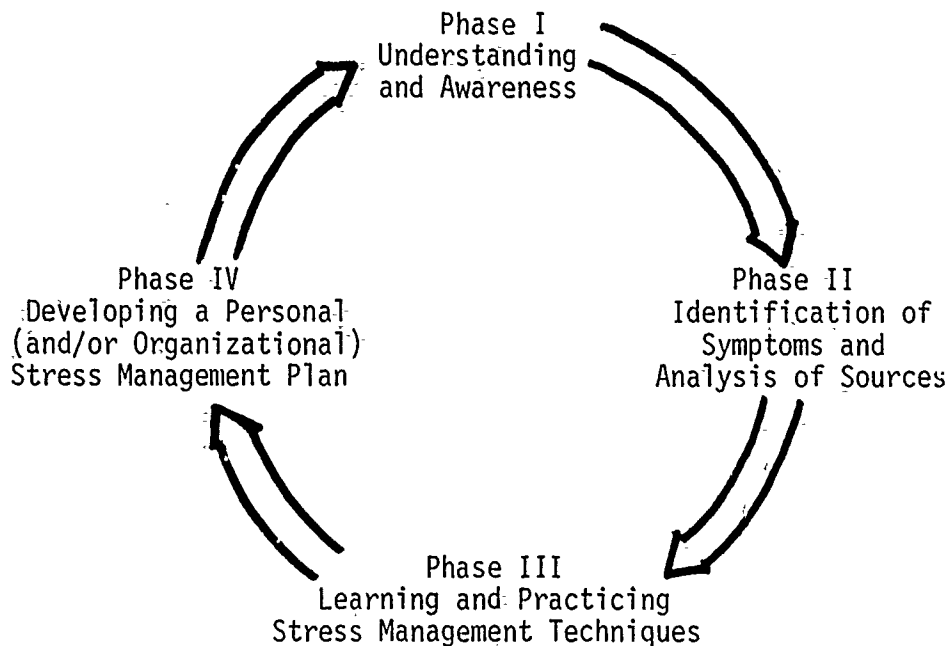


Figure 1

Four Phase Stress Management Approach

Phase I consisted primarily of lecture, supplemented by an experiential ice-breaker exercise which helped introduce the subjects of cognitions, perception and rational self-talk. Following that each participant completed four measures of stress: episodic work-related, episodic nonwork-related, chronic work-related, and chronic nonwork-related. The instruments used can be found in Adams' (1980) stress management workbook.

Although our primary interest was job-related stress we included measures of nonwork-related stress because of the obvious truism that employees do not leave home or family-related stress in the parking lot with their autos when they come to work in the morning. If one's spouse is seriously ill or if a teenager is having problems with drugs, the anxiety generated by these concerns will affect both individual and organizational effectiveness.

To the degree that individuals and their supervisors can develop skill in recognizing cues which signal a possible stress overload, they can be effective in initiating appropriate stress management action. Therefore, part of the workshop was devoted to understanding the nature of stress and recognizing its symptoms in ourselves and others.

After becoming aware of the presence of stress, participants were trained at identifying the sources of stress. Figure 2 reflects the four types of stress measured, and relates them to typical sources.

	<u>Work-Related</u>	<u>Nonwork-Related</u>
<u>Episodic</u> <u>Stress</u>	<u>Type I</u> (e.g., new boss)	<u>Type II</u> (e.g., illness in family)
<u>Chronic</u> <u>Stress</u>	<u>Type III</u> (e.g., too much to do at work, too little time)	<u>Type IV</u> (e.g., losing ground to inflation)

Figure 2

Sources of Stress

Measures of Stress

The measures of Types I and II stress are derived from the work of Holmes and Rahe (1967) and focus on the demands placed on the individual to adapt to recent events. Type I stress (episodic, work-related) is measured by summing the weights assigned to 30 recent (within the last 12 months) events. The events are adapted from the work of Naismith (1975) and range in severity from "being transferred against my will to a new position or assignment" (81 points, to "undergoing a minor relocation of my work place" (5 points). The significance of the point values is that the weight assigned to each event reflects what Adams (1978) describes as ". . . the average amount of readjustment required for one to feel 'back to normal' following that change event."

Type II stress is also measured by summing the weights assigned to events experienced within the past 12 months. Adapted from the work of Cochrane and Robertson (1973), the episodic nonwork-related events ranged from "death of a spouse" (86 points) to "development of friendship with new neighbors" (18 points). As with Type I events, the point values indicated the amount of readjustment required to "get back to normal."

As Holmes and Rahe (1967) discovered with their Social Readjustment Rating Scale (SRRS), the number of readjustment points accumulated by individuals during a twelve month period correlated significantly with the number of illnesses they experienced. This approach at measuring stress accepts Selye's view of stress related diseases as resulting from an overload of the body's ability to adjust to external demands placed upon it.

Types III and IV stress measurement was accomplished through instruments adapted from the work of French and Caplan (1972) and their colleagues at the University of Michigan. The Michigan team worked extensively with NASA in the late 60's and early 70's, studying day-to-day (chronic) stress and its effect on health and well-being.

Type III stress (chronic, work-related) is measured by applying a five-point frequency scale to 25 conditions which might be descriptive of one's work situation. These conditions include: work overload, role ambiguity or conflict, territoriality, poor-quality relationships, interunit conflict and lack of participation. The frequency with which individuals experience these kinds of conditions have been found to be significantly correlated with their feelings of competence, effectiveness, satisfaction and growth.

Type IV stress (chronic, nonwork-related) is measured by a five-point Likert stress severity scale applied to 16 potentially stressful conditions of home and social life. Scores for both Types III and IV stress are obtained by summing the responses to the 25 day-to-day stressful conditions at work and the 16 home and societal stressors.

Mean stress scores reported by the 231 DoD participants and 121 non-DoD civilian personnel attending similar workshops during the same time period were:

<u>Type of Stress</u>	<u>DoD Mean</u>	<u>Civilian Mean</u>
Episodic, work-related (Type I)	370	259
Episodic, nonwork-related (Type II)	178	181
Chronic, work-related (Type III)	62	59
Chronic, nonwork-related (Type IV)	35	33

Discussion

"Stress" has negative connotations in our society. When a manager is described as being under a great deal of stress the unstated message is that the performance of the manager in question has been slipping and is no longer up to standard. Such usage of the word reflects a rather pervasive lack of understanding about the nature of stress. Stress is very much a factor in our daily lives; however, we all manage stress differently. Further, some of us are relatively effective in our management of stress while others of us, because of a lack of stress management skills, can be classified as among the ranks of the "walking wounded."

No one seriously suggests that the elimination of stress on the job should be the goal of any organization. An essential part of any manager's job is the intentional creation and management of stress as a dynamic motivating force within the organization. The key thought is that stress should be planned and controlled: too little stress and performance will be flat and apathetic, too much stress and activity will be frenetic and similarly unproductive.

From the data provided by workshop participants, the DoD personnel appeared to be experiencing significantly higher current work-related stress than other groups of individuals administered the stress instruments. According to national data collected by Adams (1980, 23), the DoD participants ranked in the 77th percentile. An elevated level of Type I stress also is reflected in the comparison of the DoD mean with that of non-DoD workshop participants (370 vs. 259).

High levels of Type I stress reflect the presence of disruptions to the equilibrium of organization members. Although the high levels of Type I stress have been found to correlate highly with stress related disease, a number of researchers (Adams, 1980) have found that Type III (chronic work-related) stress has a more devastating impact on organization members. These findings are consistent with Selye's (1974) General Adaptation Syndrome (GAS) and the notion that the human body has a finite capacity to adapt to stress.

From a behavioral science intervention viewpoint, it is interesting to note that informal norms are the principal sources of daily on-the-job stress. When the work situation equilibrium is upset by significant organizational changes, dysfunctional work norms tend to exacerbate the situation and maintain stress at elevated levels. This view of stress in the work organization seems to fit quite nicely with process consultation and suggests that the OD consultants might be well advised to add stress management to their repertoire of behavioral interventions. In fact, a number of OD specialists now contend that work-related stress must be diagnosed and treated at the level of the individual manager before OD interventions can be effectively initiated (Suojanen and Hudson, 1981).

Lessons Learned

The initial stress management workshop was designed along the lines of the four phase model presented in Figure 1. The focus was not on the organization, but on the individual. Specifically, the objectives included understanding of the nature of stress, recognizing signs of stress in ourselves and others, identifying major sources of stress in our current lives, becoming familiar with and practicing some stress management techniques and developing a personal stress management plan.

It was not anticipated that demand would be such that nine workshops would be offered; therefore, collection of data in the early workshops was accomplished primarily for pedagogical use. But before long it became evident that an opportunity was being missed in that the data being gathered informally was potentially quite useful to senior management in the sponsoring organizations. In fact when the data were reviewed in the course of later workshops participants typically expressed the hope that the data would be provided to their senior managers.

In the end an informal summary report was provided to the commander and senior staff of the installation. Unfortunately, the report did little more than recount summary statistics supplemented with anecdotal observations. Were a similar program to be designed now, instrumentation based upon the work of French and Caplan (1972) would be added. Specifically, in addition to the four types of stress, we would focus on their ten major work-related sources of stress and use these data to help examine organizational norms. Problematic areas would then be highlighted and provide a focus for future interventions. Otherwise, the approach described in this paper was found quite workable, and seems to hold considerable promise.

REFERENCES

- Adams, J. D. "Improving stress management: an action-research-based OD intervention," from W. W. Burke, ed., The Cutting Edge (San Diego: University Associates, 1978).
- Adams, J. D. Understanding and Managing Stress: A Workbook in Changing Life Styles (San Diego: University Associates, 1980).
- Cochrane, R. and A. Robertson. "The life events inventory." Journal of Psychosomatic Research, 1973, 17, pp. 135-139.
- French, R. P. and R. D. Caplan. "Organizational stress and individual strain." In A. J. Marrow, ed., The Failure of Success (New York: AMACOM, 1972).
- Holmes, T. H. and R. H. Rahe. "The social readjustment rating scale." Journal of Psychosomatic Research, 1967, 11, pp. 213-218.
- Ivancevich, J. M. and M. T. Matteson. "Optimizing Human Resources: A Case for Preventive Health and Stress Management." Organizational Dynamics, Autumn 1980, pp. 5-25.
- Kreitner, R., S. D. Wood and G. M. Friedman. "Warning: Your Job May be Killing You." Business, Jan-Feb 1981, Vol. 31. No. 1, pp. 2-6.
- Naismith, D. "Stress among managers as a function of organizational change." Unpublished doctoral dissertation, George Washington University, 1975.
- Selye, H. Stress without Distress (New York: J. B. Lippincott, 1974).
- Suojanen, W. W. and D. R. Hudson. "Coping with Stress and Addictive Work Behavior." Business, Jan-Feb 1981, Vol. 31, No. 1, pp. 7-14.

BATTLEFIELD STRESS:
MANAGEMENT TECHNIQUES

Russell J. Hibler, Major, USAF, BSC
USAF Hospital Wiesbaden (USAFE) Germany

Abstract

A battle stress management plan that is operated at a European Air Base is reviewed. Based on mental health principles it uses assigned mental health and social actions personnel to provide line consultation and stress management triage in the three echelon medical care system. Assistance to line personnel is to provide through identifying and managing events critical to effective psychological functions. Medical screening consists of rapidly separating administrative from medical factors and in recognizing the severity of psychological responses to battle. Treatment is offered through rest, a relief group, a trial at work and, for most, a return to duty. The principles of immediacy proximity and expectancy are used throughout these procedures and serve as guidelines for care and planning.

Current literature surveys have begun to recognize the historical evidence that psychiatric combat casualties are unavoidable. There is a growing conviction that such losses, if managed properly, can be recycled back into the manpower pool (Ingraham and Manning, 1980). However, until recently those personnel who might be able to assist with psychiatric casualties have been assigned alternate wartime duties. This article describes a battlefield stress management plan that is based on mental health principles that have been developed and tested since October 1980 at Bitburg Air Base, Germany. The rationale, structure and functioning of the plan are described, rather than emphasizing the theoretical and historical antecedents. Mental health and social actions personnel provide the management, and this plan explains how they fulfill their roles and where they function in the three echelon medical system.

This plan uses mental health and social actions personnel as a team that provides the initial principles of stress management on the battlefield. During peacetime exercises and program development the stress management team includes the clinical psychologist, clinical social worker, psychiatrist, mental health technicians, and social actions personnel. During wartime, it is envisioned that the psychiatrist would be required to function primarily as a physician and would occasionally be available as a consultant to the team. Because of this assumed limited psychiatric availability, the senior BSC officer was identified as the team chief. The mental health technicians and social actions personnel were assigned in equal numbers to the psychologist and the social worker. As much as possible, these two groups are matched for experience and training qualifications. Each of these teams comprised an independently functioning unit that was capable of deployment and, with augmentees from the manpower pool, was able to provide 24 hour coverage.

Although the standard mental health education only partially prepares these personnel for wartime triage and consultation, additional academics and OJT is readily available. We found that surveying the literature would provide a working knowledge of the subject matter within several days.

The references cited in this report both review the relevant empirical literature and provide a comprehensive overview of the Army's current status in this area. Additionally, non-traditional experiences such as simulation in the field have also been invaluable to helping us understand the wartime medical environment and the battlefield. Frequent deployment with air transportable hospitals and clinics, and second echelon treatment teams provided a realistic awareness of the working environment at several locations and under all weather conditions. This "hands on" experience also allowed the development of simple and practical triage techniques.

At least in Europe, Air Force medical facilities provide direct support for large numbers of Army units. We felt that in order to serve these organizations, as well as to understand the etiology of wartime casualties, it is necessary to participate in Army field exercises. This experience included maneuvers with mechanized infantry that was enlightening but to some degree anticipated through past contacts with war stories, movies, etc. What was not anticipated, yet vividly revealing, were the less heroic battlefield aspects including the hours of boredom followed by pandemonium, having mechanical failures, becoming lost, engaging your own personnel and losing troops to drug abuse. Overall, the complexity and newness of this experience shattered the hope of being able to anticipate all the stress management needs for an organization, but it did provide a pragmatic and realistic framework for understanding the normal range of functioning on the battlefield. This was particularly important since "before anyone can understand, prevent, manage and restore to effectiveness a psychiatric casualty, one must have a firm sense of the normal reactions of soldiers to combat" (Rath, 1980, p.13). With the battle stress management team identified and their training outlined, their provider roles can be described next.

Management Roles: Consultation to the Line

As line consultants, the battlefield stress management team works to enhance ongoing leadership and existing management techniques. This is done by evaluating group and environmental factors that determine the incidence and types of combat ineffectiveness. Many of these factors require a working knowledge of the organization being served as well as organizations in general. The ongoing familiarization outlined above provides an awareness of the specific units. The more general organizational factors include: "(a) intensity of combat, (b) duration of combat, (c) type of combat action, (d) pace of combat action, (e) wounded in action and killed in action rates, (f) type of unit, (g) unit cohesion, (h) unit leadership, (i) replacement process, (j) experience in combat, (k) expectations, (l) competing demands on loyalty, (m) command preparation for management, and (n) medical preparation for management" (Rath, 1980, p.14,a). These factors have a variety of uses including the epidemiological prediction of casualties, and the estimation of the timing when psychologically based problems begin to occur.

Manning (1980) provided one example of such consultation services on the relationship between fatigue and performance. He found that round-the-clock military exercises resulted in great impairment in psychological or cognitive process skills, while less errors occurred in skills that were more psychological in nature. In fact, despite the environmental stresses and lack of sleep, "the well-learned, and mainly physical tasks were highly resistant to deterioration" (page 10).

He also found that for demanding situations that required cognitive skills, those in charge often chose leadership styles founded on willpower and conspicuous self-denial. The result was a high susceptibility to fatigue and a decrease in the quality of the leader's work.

After only one night without sleep, task orientation deteriorated with decreased to less than 50% of the normal performance levels. A similarly demanding lifestyle (minimal rest, food, etc.) on the part of more physical laborers would only decrease the quantity of work. Other results showed that including naps as part of effective leadership behavior, as well as scheduling a minimum of as little as three to four hours of sleep per night yielded a significant performance improvement (p.11). In short, identifying such critical behaviors, providing relevant information and supplying personnel support may provide organizations as well as individuals with increased resistance to fatigue and increased performance.

Management Roles: Medical Screening

The second and more familiar consultant role for the stress management team is the screening and psychological triage of casualties. In essence, the team would identify normal somatic and psychological reactions to battlefield stresses and screen out those who have abnormal responses. The normal reactions include "(a) muscular tension, (b) shaking and tremor, (c) increased perspiration, (d) digestive and urinary system reaction, (e) circulatory and respiratory system reactions; and normal psychological reactions included: (a) fear and panic, (b) sensitivity to noise, (c) sleep difficulties, (c) pathetic tendencies, (e) irritability and resentment, (f) extremely lethargic or euphoric post combat mood states" (Rath, 1980, p.14,a). One or more of these normal reactions occur in most personnel and therefore their presence should be recognized and legitimized. Most people can tolerate these symptoms if they understand and expect them and they do not need additional intervention or support. These "normal" reactions are separated from "abnormal" reactions by a continuum based on the severity, number and duration of these symptoms and their effect on the individual's ability to perform his duty.

Three familiar psychiatric reactions that are often experienced in peace time populations are also seen on the battlefield. The first psychiatric reaction is overt psychoses. The symptoms of overt psychoses are very visible and result in a low return to duty rate. They are expected to be easy to identify since their symptoms frequently become apparent soon after the onset of hostilities. A second familiar psychiatric reaction is the onset of psychosomatic syndromes. These physiological symptoms are less obvious and most frequently involve gastro-intestinal syndromes, low back pain, or headaches that do not present organic findings. Additionally, casualties result from failure to take proper precautions. The result and injuries are therefore

partially psychological in nature. These include accidents, and illnesses such as frostbite that could have been prevented by appropriate prophylactic measures. The third type of syndrome is the psychologic complications of medical problems. In effect, medical recovery becomes slower than expected, there is increased verbal or behavioral complaints and these conditions are often largely ignored by the providers. Both these psychosomatic and psychologic complications are readily managed by providers when there is a clear expectation that the casualties will return to duty.

The remaining four syndromes are unique to the battlefield environment. The first and most frequent response is the combat reaction. Overall, this is an acute anxiety reaction manifested by an exaggeration of the normal combat reaction with an accompanying feeling of impending death or maiming. The net result is that the individual is unable to perform his duties. This response occurs very early in combat and is closely associated with the intensity and type of stressors. Further decompensation is prevented if handled by a prompt return to duty. Otherwise, these normally transient conditions become more chronic and serious (Levan, Greenfield, and Baruch, 1979).

A second type of combat response is a withdrawal from battle. This refusal to fight is due to a variety of motivational or societal issues. Included are being AWOL, alcohol or drug abuse, or being preoccupied with dependents. This response may occur early after entering combat but usually develops in large numbers as time progresses. It is handled best by line personnel through phone calls home, rest and recreation trips, increasing BX services, etc. Similarly, combat exhaustion, a third combat syndrome, results from the stress of prolonged combat of entire units. It is best handled through a group rotation to noncombat duties. A fourth combat response is the short timers syndrome. This is effected by organizational rules, group names, and the expectations of the individual. Basically, as the individual gets "short" his anxiety increases and he becomes reluctant to expose himself to danger. Again, good modeling by leaders and organizational remedies rather than medical actions are indicated. Finally, as the types of war actions change, additional syndromes occur. The team must be vigilant to detect new responses and be judicious in providing assistance.

Diagnoses may even damage the casualties' prognosis. Danish and Smyer (1981) reviewed the helplessness literature and concluded that diagnosing and labeling individuals decreases their perceived control and may intensify their symptomatology. Additionally, Noy (1980) found that reactions of even psychotic proportions responded favorably if treated immediately and firmly without a diagnosis. For these reasons our team uses only the syndromes or battle stress types:

Battle shock. The immediate or acute onset of emotional and somatic signs in which severe anxiety symptoms predominate and last up to 72 hours. Ideally, this is managed at the lowest level of care possible. Only those who present with most severe and life endangering behaviors are evacuated for additional care.

Battle fatigue. These symptoms are similar to battle shock but they exist with an inability to perform duties for over 72 hours. These symptoms may be more chronic and insidious in their onset. Evacuation to nearby medical facilities is indicated only after initial trials at sleep, rest, etc. are ineffective.

Treatment Priorities

The principles of immediacy, proximity, and expectancy for providing care are repeatedly cited in the literature as essential criteria (Levan, et al, 1979). This literally means treating battle stress casualties within minutes or hours after the onset of the symptoms. It also means providing food, warmth, rest, and relative safety or other support as proximal as possible to their duty section. This distance is often psychological rather than geographical resulting in "felt" safety that can actually be as close as one hundred yards from the action. Finally, and perhaps most important, is the expectancy that they will return to duty. This expectancy must be expressed at all levels of casualty management. By comparison, if casualties are treated as "patients" and evacuated to the rear, they must frequently maintain their symptoms longer, rarely return to the same duty, or they may not return at all. By using these policies of immediacy, proximity, and expectancy most casualties are returned to duty in less than 72 hours. They also provide the basis for the management techniques at each level or echelon of care.

The Three Echelon Medical Care System

The three echelon medical care system was developed to provide a flexible medical delivery system that can operate in environments that are free of combat or of nuclear, biological, or chemical warfare contaminants. Within this system the first echelon of care is provided by "buddy care". Although somewhat limited in scope, these first aid skills are immediate and are locally available. Individuals who do not respond and require more definitive medical attention are marshalled at an on-base shelter that is designated the casualty collection point. From there they are transported off-base to the second echelon of care. This secondary care is located in tents or other available shelter approximately two to four miles from base. The area is selected for being relatively safe from hostilities or contamination. At this site, the casualties are decontaminated as necessary and have their first contact with medical personnel. They are triaged by the severity

of their medical condition given further evaluation and their condition stabilized in immediate, delayed and minimal treatment areas. These services provided either minimal care and a return to duty within several hours, or immediate and delayed care that may result in transportation to the third echelon. This third level of care is located in a safer area approximately twenty miles away, in more permanent or hardened facilities and has ample staff and resources to provide longer term and definitive medical care. Following the treatment, the patients are either directly returned to duty or prepared for medical evacuation to other overseas or CONUS facilities for long term care.

Battle Stress Management Locations

Initially, individuals must manage their battle shock on their own, with the help of "buddy care", or receive support and reassurance through their chain of command. If these first echelon measures are insufficient, they are referred through the casualty collection point to the second echelon where the battle stress management personnel work in close physical proximity to the minimal care treatment team.

This placement is essential since many of the battle stress casualties may have concurrent somatic injuries. Additionally, since their treatment priority is minimal, the close functioning with the minimal treatment simplifies the management process and reinforces the expectations of staff and casualties alike that they will return to duty.

Several forms of abbreviated mental status checklists have been used by the battlefield stress team and can provide brief screening information in less than 10 minutes. Those who have not eaten, or slept recently are given a meal and a sleeping space for several hours prior to their return to duty. Those individuals who refuse to fight or have difficulties, are referred back to their unit for organizational management. Casualties who have the more chronic battle fatigue or very severe battle shock and would be transferred for additional medical management may receive neuroleptic medications to initiate rest. The medical personnel on the minimal team maintain and administer the psychiatric medications with consultation from the battle stress team. The consultant or second function of the battle stress management team is often provided during the lulls between triage and direct casualty management. Radio communications and house calls, as necessary, are readily available to provide feedback, support, and other information to the personnel on base. The team also functions as a trouble shooter for the stress that the medical personnel receive. This is important to prevent the rescue team from becoming casualties (Martin, 1980).

At third echelon, another battle stress team functions similarly to the team at the second echelon site. It was also prepared to initiate the following treatment regimen for a one to four day period:

1. Rest. During this phase, hot food, warmth, safety, support and sleeping space (preferably not a bed) are provided. If sleep does not occur naturally after a meal, it may be induced as needed. This initial phase provides psycho-physiological support to counteract exhaustion. Wherever possible, the casualties' condition is referred to as battle fatigue and they are assured that they will return to duty. Rather than use pajamas, they remain in their uniforms at all times to preserve their role identity. Additionally, the psychiatrist will function at this facility and, as available, provide consultation services to the team.

2. Relief group. Morning group sessions are used to share grief, anger, fear, etc., and to receive both directive and supportive counseling from the staff. This process serves to address and legitimize built up feelings and somatic responses, and to recognize their coping skills.

3. Recycle. As the casualties become rested, they supplement the manpower and minimal treatment teams by being assigned appropriate tasks. This work is in itself therapeutic (Chapman, Hackley, and Spencer, 1980). Their ability to perform these tasks also tests their readiness to return to duty.

4. Return to duty. When steps one through three have been successfully completed, the casualties are returned to their unit. If significant pathology persists, they will be considered for evacuation to other, more distant, treatment facilities, or to the CONUS facilities. The battle stress management plan is summarized below. Additions to the plan are being considered to extend its potential usefulness.

Discussion

Perhaps even more important than the details of this plan is its ability to recognize and legitimize psychological reactions to battle, and the commitment of the plan to the use of available resources to provide enlightened management. The specifics of the plan provide for flexibility with who, where, and when assistance is rendered. The operational principles of the battle stress management team at Bitburg Air Base, Germany, involve the combination of mental health and social actions personnel in roles that provide direct casualty care and use expertise in stress management. Additional OJT and less traditional training is necessary to perform these roles.

Training includes field work and familiarization with battle field conditions. Battle stress management is provided by personnel through consultation to line units and by screening casualties for extreme or long standing psychogenic reactions that effect their duty performance. They provide a triaging of such casualties and initiate management procedures that include food, shelter, ventilation of feelings, and a trial at returning to work prior to being returned to duty. Throughout, the principles of immediacy, proximity, and expectancy provide guidelines for overall delivery of services. By consideration of these ideas for use in peacetime planning, the responses of battlefield casualties and their organizations may be more adaptive and health providing.

References

- Chipman, M. Hackley, B.E., and Spencer, T.S. Triage of mass casualties: Concepts for coping with mixed battlefield injuries, Military Medicine, 1980, 145, No 2, 99-100.
- Danish, S.J., and Smyer, M.A. Unintended Consequences of Requiring a license to help. American Psychologist, 1981, 36, 13-21.
- Ingraham, L.H. and Manning, F.J. Psychiatric Battle Casualties: the missing column in a war without replacements. Medical Bulletin of the U.S. Army Europe, 1980, 37, No. 12, 3-9.
- Levan, I., Greenfield, H. and Baruch, E. Psychiatric combat reactions during the Yom Kippur War, American Journal of Psychiatry, 1979, 136, No.5, 637-641.
- Manning, F.J. Continuous operations in Europe: Feasibility and the effects of leadership and training, Paramedics, Journal of the U.S. Army War College, 1980, 9, No 3, 8-17.
- Martin, R.D. Disaster planning for psychiatric casualties for the general hospitals with psychiatric services, Military Medicine, 1980, 145, No 2, 111-113.
- Noy, S. Modern Warfare: The Israeli Experience. Paper presented at the fifteenth Anglo-American Military Psychology Symposium, London, October 1980.
- Rath, F.H. The psychiatric casualty and the combat effectiveness-ineffectiveness continuation, Medical Bulletin of the U.S. Army Europe, 1980, 37, No 11, 3-9. (a)
- Rath, F.H. Estimating Psychiatric Casualty Workload. Paper presented at the 1980 U.S. Army Europe and Seventh Army Annual Medical Surgical Conference, Garmisch, Germany, 1980. (b)

WHAT EXPERIENCE TEACHES:
COGNITIVE RESPONSES TO A STRESSFUL ENVIRONMENT

Earl H. Potter, III
U.S. Coast Guard Academy

ABSTRACT

Previous studies have found that experience as measured by time in service or time on the job contributes to improved job performance. This study explores the process through which experience and experience interactions with stress and cognitive ability contribute to the academic performance of cadets at the U.S. Coast Guard Academy. The results suggest that a significant portion of the contribution of experience to improved cadet performance is due to cognitive factors, e.g., beliefs, attitudes, and concerns which change as a function of time spent as a cadet at the Academy.

The United States Coast Guard, as do other civilian and military organizations, relies on job experience to teach many of the skills necessary for effective leadership and technical performance. Yet, the research literature is notably non-conclusive with respect to the role of experience as a teacher. For example, Bray (1966) in his original report of AT&T's assessment center technique for evaluating management potential found little change in the behavior of managers in simulated job situations after eight years of management experience. On the other hand, other studies have demonstrated that military personnel with more experience respond differently under stressful conditions (Berkum, 1964).

This study continues a series of studies which has examined the impact of stress on job performance. In two recent studies (Potter & Fiedler, 1980; Barnes, Potter & Fiedler, 1981) stress in significant interpersonal relationships was associated with a decrease in the relationship between intelligence and job performance. The size of this decrease, however, varied with job experience such that the relationship between intelligence and performance under conditions of high interpersonal stress was greater for persons with more experience in the job. In the first of these studies, Potter and Fiedler (1980) found that junior Coast Guard staff officers who reported high levels of stress with their bosses had a low correlation between their intelligence and job performance. The relationship between the intelligence and performance of more senior staff officers did not decrease with high levels of stress. In a different setting, Barnes, Potter and Fiedler (1981) recently found that experience as a cadet at the U.S. Coast Guard Academy likewise diminished the impact of stress on the relationship between intelligence (SAT math scores) and performance (GPA).

While these findings may be of some passing practical interest, they have little theoretical value because they say nothing about how experience might mitigate the impact of stress on performance. One is left wondering what changes occur as a function of time that might influence the relationships between stress, intelligence, and performance. One hypothesis has emerged from conversations with cadets at the Coast Guard Academy and Sarason's (Sarason, Johnson, Berberich, and Siegal, 1978; Sarason and Sarason, 1979) work on cognitive moderators of stress.

If we follow conventional wisdom, graduates of a military Academy should be observed to meet stressful situations with a sense of confidence and mastery derived from four years of coping with a challenging environment. In psychological parlance, the fact of their survival leads to attributions of competence and courage (or as the

cadets have told me "after four years you learn not to sweat it"). Sarason and Sarason (1979) and Meichenbaum (1975) have suggested that stress causes performance problems because anxious persons agonize and fret over the consequences of failure or danger and consequently do not focus on the task at hand. One explanation for a cadet's improved response to stress may be that cadets develop a belief in their own ability to meet challenges, and therefore, do not worry or catastrophize as much. If this is the case, then the interaction of experience with intelligence and stress may be accounted for by measurable change in cognitive responses to the stressful environment of the military Academy.

METHOD

Subjects. Two hundred thirty-four cadets at the U.S. Coast Guard Academy (73 freshmen, 67 sophomores, 30 juniors, and 64 seniors) responded to a questionnaire which included background information, personality scales and assessments of their environment. Of the subjects, most were non-minority (209) and male (207). As is the case with cadets at all academies, their academic and social history was universally successful.

Performance. Semester grade point average (GPA) was used as the indicator of performance. Typically the relationship between intelligence and any one semester is lower than the relationship between intelligence and overall GPA due to variations in course difficulty over semesters. However, stress also varies from semester to semester, and stress measured in one semester is more likely to influence performance in that semester than overall GPA. For these subjects GPA ranged from 0.70 to 3.97 with a mean GPA of 2.72.

Intelligence. The Scholastic Aptitude Test (SAT) mathematics score was used as the measure of intelligence. This score was chosen over the verbal score due to the fact that the curriculum is heavily technical in nature. The range of SATM scores was from 470 to 790 with a mean of 643. These figures represent a restriction in range which would be expected to result in an underestimation of SATM/Performance correlations.

Stress. Interpersonal stress in significant relationships was measured with a twenty-item, Likert-type scale which asked cadets to consider how much stress they felt with, for example, their instructors as a result of their instructors placing high demands for performance on the cadets while providing insufficient direction and support. A principal component factor analysis with varimax rotation found five orthogonal factors in the scale--relationship to peers, instructors, parents, company officers, and cadets senior in the chain of command. For the purpose of this study, interpersonal stress was measured by two of these factors, peer stress and instructor stress, which reflected different sources of stress and were likely to be differentially related to academic performance. Peer stress with a possible range of 4 to 28 ranged from 4 to 28 with a mean of 12.22. This score was related (.31, $p < .001$) to a measure of social evaluation apprehension (Watson & Friend, 1969). Instructor stress, with the same possible range, ranged from 4 to 27 with a mean of 11.72 and was significantly related (.29, $p < .001$) to a measure of test anxiety (Sarason, 1979).

Experience. The cadets' year group, classes of 1981, 1982, 1983, and 1984, served as the measure of experience. Few cadets had prior military experience (9) and few had prior college (30). Even if a larger number did have significantly greater prior experience, it is not at all clear that this experience would be relevant to performance at the Coast Guard Academy.

Cognitive Factors. A 23-item, Likert-type questionnaire which reflected attitudes about and beliefs concerning the challenges faced by cadets was developed as a result of numerous conversations with cadets. A principal component factor analysis with varimax rotation resulted in three factors with eigenvalues greater than 1.0 which accounted for 73.9% of the scale's variance. The first of these factors, labeled concern for approval accounted for 43.4% of the variance and included items such as:

How important is it to you that your classmates approve of you? (Extremely - Not at all)

It is worth the extra effort it requires to bring up your military aptitude grade. (Strongly Agree - Strongly Disagree)

This factor which included five items with factor loadings from .52 to .77 had a possible range of 5 to 35. Actual scores ranged from 7 to 35 with a mean of 22.76.

The second factor was labeled "doubt in my ability to meet demands." This factor which accounted for 17.2% of the scale variance included items such as:

There just isn't enough time to do both my course work and satisfy the demands of cadets senior to me. (Strongly Agree - Strongly Disagree)

Sometimes when things get rough, I give up trying to do everything at once and do nothing. (Strongly Agree - Strongly Disagree)

This factor included five items with loadings ranging from .45 to .66. Actual scores ranged from 6 to 34 with a possible range of 5 to 35. The mean score was 19.35.

The third factor contained only two items which indicated that cadets set aside studies in order to meet military requirements. Scores on this factor labeled "misdirection of efforts" ranged from 2 to 13 with a mean of 5.73.

Data Analysis. Changes in cognitive responses to stress were first assessed with a one-way analysis of variance across class groups. Next paired t-tests were performed for cognitive factors taken for 40 members of the class of 1984 at the end of their freshman year and six months later in the middle of their sophomore year.

Cognitive factors were then included in two hierarchical multiple regression analyses--one for peer stress and one for instructor stress. Main effects were entered first, multiplicative two-way interactions second and the three-way multiplicative interactions last. The experience, stress, intelligence relationship was assessed in similar fashion using two hierarchical regressions with class entered as a dummy coded co-variate in main effects, two- and three-way interactions.

The ability of cognitive factors to explain the variance in performance accounted for by experience was evaluated with a technique called communality analysis by Kerlinger and Pedhazur (1973). Communality analysis allows the statistician to estimate the amount of variance accounted for in, for example, performance by two factors independently and in common.

RESULTS

Changes due to experience. Table 1 shows that concern for approval, self-doubt, and misdirection of efforts all decrease as cadets gain in experience at the Academy. This analysis, however, is subject to the criticism that the findings may result from

Table 1

Oneway ANOVA: Cognitive factors by class

Means by class

	Freshman		Sophomore		Junior		Senior	<u>F</u>	<u>p</u>
Concern for Approval	25.32	>	23.10	>	19.97	=	20.53	10.93	< .001
Self-doubt	22.93	>	17.49	=	16.80	=	18.28	14.16	< .001
Misdirection of Efforts	8.89	>	4.33	=	4.14	=	4.33	47.50	< .001

attrition which leads to the average survivor being more confident rather than real change in each "survivor" due to his own experience. Therefore, a limited followup was conducted utilizing 40 cadets who were freshmen at the time of initial testing. These 40 cadets were given the same questionnaire six months later. Table 2 shows that even in a short time some positive change occurs. Both concern for approval and misdirection of efforts are significantly lower ($t = -2.98$, $p = .018$, and $t = -4.15$, $p .001$ respectively). There is a slight and non-significant increase in

Table 2

T-Tests for Cognitive Variables for 40 cadets
polled in their freshman (1) and sophomore (2) year

		<u>\bar{x}</u>	<u>t</u>	<u>2-tail Probability</u>
Concern for Approval	1	25.83		
	2	22.85	-2.98	.018
Self-doubt	1	20.73		
	2	22.00	1.27	.159
Misdirection of Efforts	1	10.53		
	2	6.38	-4.15	.000

self-doubt. Together these findings support the hypothesis that one consequence of experience as a cadet is less worry and self-preoccupation in the face of threat.

Experience and stress. Table 3 supports the findings of previous studies in that stress does not contribute directly to GPA nor does the two-way interaction with intelligence account for a significant proportion of the variance in GPA. There is in both cases, however, a significant three-way interaction between stress, SATM, and experience (i.e. class). Multiple R's for both instructor stress and peer stress are highly significant and comparable (.5329 and .5346 respectively).

Table 3

Summary Table: Multiple Regressions on GPA
with stress, intelligence and class

<u>INSTRUCTOR STRESS</u>					
<u>Predictor</u>	<u>Increments of R²</u>	<u>Simple R</u>	<u>Multiple R</u>	<u>R²</u>	<u>F</u>
SATM	.1360***	.37			
Instructor stress	.0057	-.12			
Stress x SATM	.0141	-.04			
Class	.0862***				
Stress x class	.0006				
Stress x SATM x class	.0414***				
			.5329	.28	5.42***
<u>PEER STRESS</u>					
SATM	.1360***	.37			
Peer stress	.0039	-.06			
Stress x SATM	.0103	.02			
Class	.0827***				
Stress x class	.0056				
Stress x SATM x class	.0473***				
			.5346	.29	5.47***

*** $p < .001$

Cognitive factors and stress. Table 4 summarizes the hierarchical multiple regressions which include stress, intelligence, and three cognitive factors. Taken together, cognitive factors and their interactions with instructor stress and intelligence account for 10.1% ($p < .025$) of the variance in GPA. Instructor stress is not directly related to GPA and SATM alone accounts for 13.6% of the variance in GPA. The relationship of stress with the instructor to GPA varies with self-doubt ($F = 10.07$, $p < .005$) and concern for approval alone accounts for a significant amount of the variance of GPA ($F = 4.69$, $p < .05$). The multiple R is .5002 which is highly significant ($F = 4.71$, $p < .001$).

The second equation shows the interactions with peer stress, intelligence, and the cognitive factors. While the multiple R for this equation (.4630, $F = 3.73$, $p < .001$) is still highly significant, less variance for GPA is accounted for by cognitive factors and their respective interactions with peer stress and intelligence

Table 4

Summary Table: Multiple Regressions on GPA
with stress, intelligence and three cognitive factors

INSTRUCTOR STRESS

Predictor	Increments of R^2	Simple R	Multiple R	R^2	F
SATM	.1360***	.37			
Instructor stress	.0057	-.12			
Stress x SATM	.0141	-.04			
Concern for Approval	.0213*	-.23			
Self-Doubt	.0091	-.19			
Misdirection	.0075	-.13			
Stress x Concern	.0028	-.19			
Stress x Doubt	.0457***	-.23			
Stress x Misdirection	.0014	-.18			
Stress x Concern x SAT	.0091	-.12			
Stress x Doubt x SAT	.0001	-.17			
Stress x Misdirect. x SAT	.0036	-.13			
			.5062	.26	4.71***

PEER STRESS

SATM	.1360***	.37			
Peerstress	.0039	-.06			
Stress x SAT	.0103	.02			
Concern	.0183	-.23			
Self-doubt	.0153	-.19			
Misdirection	.0059	-.13			
Stress x Concern	.0000	-.16			
Stress x Doubt	.0004	-.13			
Stress x Misdirection	.0004	-.12			
Stress x Concern x SAT	.0207*	-.08			
Stress x Doubt x SAT	.0027	-.06			
Stress x Misdirect. x SAT	.0004	-.08			
			.4630	.21	3.73***

* $p < .05$

*** $p < .001$

(6.41%, $p = .15$). None of the main effects for cognitive factors or two-way interactions are significant. One three-way interaction (stress x concern for approval x SAT) is significant ($F = 4.27$, $p < .05$). As hypothesized, cognitive responses to stress which vary with experience are psychologically relevant to the impact of stress on academic performance.

represents the total variance and the main effects and the stress may be added to class in successive steps. The increase in the R^2 of the instructor stress equation

Table 5

Summary Table: Multiple Regressions on GPA
with stress, intelligence, cognitive factors and class

INSTRUCTOR STRESS

Predictor	Increments of R^2	Simple R	Multiple R	R^2	F
SATM	.1360***	.37			
Instructor stress	.0057	-.12			
Stress x SATM	.0141	-.04			
Concern for Approval	.0213*	-.23			
Self-Doubt	.0091	-.19			
Misdirection	.0075	-.13			
Stress x Concern	.0028	-.19			
Stress x Doubt	.0457***	-.23			
Stress x Misdirection	.0014	-.18			
Stress x Concern x SAT	.0091	-.12			
Stress x Doubt x SAT	.0001	-.17			
Stress x Misdirect. x SAT	.0036	-.13			
Class	.0592***				
Class x Stress	.0029				
Class x Stress x SATM	.0192				
			.5810	.34	3.76***

PEER STRESS

SATM	.1360***				
Peerstress	.0039				
Stress x SAT	.0103				
Concern	.0183				
Self-doubt	.0153				
Misdirection	.0059				
Stress x Concern	.0000				
Stress x Doubt	.0004				
Stress x Misdirection	.0004				
Stress x Concern x SAT	.0207*				
Stress x Doubt x SAT	.0027				
Stress x Misdirect. x SAT	.0004				
Class	.0532***				
Class x Stress	.0092				
Class x Stress x SATM	.0367*				
			.5598	.31	3.37***

* $p < .05$

*** $p < .001$

from .2563 to .3375 (.0812) is therefore the variance uniquely attributable to class. Subtracting this value from the contribution of class to the R^2 for the instructor stress equation in Table 3 (.1282) we find that class and cognitive factors in common account for 4.70% of the variance in GPA. In Table 5 we see that the total variance accounted for in GPA by class and cognitive factors is .1817. Solving the following equation (where U = unique variance and C = common variance) we find that the unique contribution of cognitive factors to GPA is:

$$R^2_{GPA}(\text{cognitive factors})(\text{class}) = U(\text{Cognitive} + U(\text{Class}) + C(\text{Cognitive} \cdot \text{Class}))$$

$$.1817 = U(\text{Cognitive}) + .0812 + .0470$$

$$U(\text{Cognitive}) = .0535$$

Likewise, the same steps for peer stress lead to the following unique and common variances:

$$R^2_{GPA}(\text{Cognitive})(\text{Class}) = U(\text{Cognitive}) + U(\text{Class}) + C(\text{Cognitive} \cdot \text{Class})$$

$$.1631 = .0275 + .0990 + .0366$$

These findings suggest that cognitive responses to stress do indeed account for a portion of the variance in GPA explained by experience. Cognitive responses appear to play a larger role in the mediation of instructor stress than in the mediation of peer stress. Cognitive factors do not, however, account for all the variance attributable to experience. Both experience and the cognitive factors included in this analysis contribute some unique portion of variance to the explanation of GPA.

DISCUSSION

This study offers support for the hypothesis that experience results in changes in cognitive responses to stress which mediate the impact of stress on academic performance. That these findings are significant but not dramatic merely attests to the complexity of the changes which result from four years of experience at a military Academy. Further refinements in the measure of change and in inclusion of other variables which change due to experience should result in a clearer picture of how experience benefits performance in this one setting. It is not at all certain that these findings will generalize to other settings or that experience gained in one setting is transferable to a different task or environment.

Practically speaking, these findings suggest the possibility of interventions to improve confidence and decrease self-doubt. Theoretically these findings offer a link between the work of Fiedler and Sarason in that the effects observed by Fiedler can now be related to the theory proposed by Sarason.

Bibliography

- Barnes, V., Potter, E.H. & Fiedler, F.E. The effect of interpersonal stress on the prediction of academic performance, Unpublished paper, 1981.
- Berkum, M.M. Performance decrements under psychological stress. Human Factors, 1964, 88, 24-26.
- Bray, D.W. & Grant, D.L. The assessment center in the measurement of potential for business management. Psychological Monographs, 1966, 80 (17, Whole No. 625).
- Meichenbaum, D.H. A self instructional approach to stress management: A proposal for stress inoculation training. In C.D. Spielberger and I.G. Sarason (Eds.). Stress and Anxiety, Vol. 1, New York: Halsted Press, 1975, pp. 237-259.
- Kerlinger, F.N., & Pedhauzer, E.J. Multiple Regression in Behavioral Research. New York: Holt, Rinehart & Winston, Inc., 1973.
- Potter, E.H. & Fiedler, F.E. The utilization of staff member intelligence and experience under high and low stress. Academy of Management Journal, 1981, 24, 361-376.
- Sarason, I.G., Johnson, J.H., Berberich, J.P. and Siegel, J.M. Helping police officers cope with stress: A cognitive-behavioral approach. Office of Naval Research Technical Report No. SCS-CS-005. Seattle, Washington, 1978.
- Sarason, I.G. & Sarason, B.R. The importance of cognition and moderator variables in stress. Office of Naval Research Technical Report No. SCS-IS-009. Seattle, Washington, 1979.
- Watson, D. & Friend, R. Measurement of social-evaluative anxiety. Journal of Consulting and Clinical Psychology, 1969, 33, 448-457.

Training for Combat Stress Reactions

A. David Mangelsdorff and T. Paul Furukawa
Health Care Studies Division, Academy of Health Sciences
Fort Sam Houston, Texas 78234

Abstract

Combat produces stress reactions. Similar reactions in most other situations would be considered abnormal responses; in the context of battle, combat stress reactions are normal responses to an abnormal situation. Commanders and their units need to prepare for combat stress; prevention and recognition of combat stress reactions can be accomplished through realistic training. Mental health professionals are integral to the planning and development of training packages for coping with combat stress.

Combat stress reactions have been recognized since the Kumano War in Japan back in 603 B.C. Ancient records reflect that the soldiers seemed to become ill and deteriorate with prolonged campaigns; the malaise disappeared after the troops were rested and morale was restored. In the American Civil War, combat stress reactions took the form of "insanity", "drunkenness" and "nostalgia". Over 5,000 cases were hospitalized for "nostalgia," a condition characterized by "mild insanity caused by disappointment and longing for home." In the Russo-Japanese War (1904 to 1906), the first accurate descriptions of combat stress reactions were made. Forward treatment centers were established and proved effective. Before the American Expeditionary Force went to Europe in World War I, a committee was sent to France to observe the treatment of casualties. A natural field experiment in treatment modalities had occurred. The British evacuated their psychiatric casualties back to the United Kingdom, while the French treated their stress reactions close to the battle lines. By treating the troops close to the fighting and by maintaining the casualties as soldiers rather than as patients, the French returned about 90% of their psychiatric casualties to combat or combat support duties. When the American Forces entered the war, a psychiatric hospital was established close to the lines. The effectiveness of forward area treatment in World War I ranged from 40 to 70% of the combat psychiatric casualties being returned to some type of duty in forward areas. In World War II, the lessons learned from World War I were not remembered. No effective treatment/evacuation policy existed until 1943. With an effective treatment/evacuation policy in World War II, for every 100 psychiatric casualties; from 40 to 60 were returned to full combat duty, while 20 to 40 were returned to some non-combat duty.

There are a number of factors which appear to place a soldier at risk of a stress reaction during armed conflict. Troops that have just recently entered the service are greater risks than those with two or more years of military service. A previous history of emotional instability before entering the service might predispose a soldier to be more susceptible to stress reactions. Troops in positions where the battle lines are static and troops that are subjected to continual bombardment without retaliation are more likely to become stress casualties. The breaking point of the average soldier in Europe during World War II was in the range of 80 to 90 days of continual combat. Fatigue and exhaustion are significant factors in combat stress reactions. Support units are more vulnerable because of not being able to perceive the usefulness of their work efforts. Group characteristics of: low morale, poor leadership, loss of a leader (particularly for psychiatric reasons), lack of

purpose, low group cohesiveness, lack of training or equipment, all contribute to increasing the risk of stress reactions.

Attempts at making comparisons between years for incidence of admission rates for psychiatric conditions are limited. Diagnostic categories change, new labels are employed, different types of groupings are made in each conflict. The recorded statistics only reflect what was recorded, not necessarily what happened. Many stress reactions may have not been admitted into the medical channels. Admission rates for psychiatric disorders vary directly with the proportion of new members in the service. With increases in Wounded in Action casualties (WIAs), there are parallel increases in the number of psychoneuroses reported; stress reactions are direct functions of the intensity, nature, and duration of the conflict.

Combat stress reactions (a soldier's means of coping with conflict) result in non-effective behaviors. Combat stress reactions are casualties in that the soldier is physically and mentally unable to perform his duty. The intensity, nature, and duration of combat influence the number of stress reactions. In World War II in infantry battalions, neuropsychiatric admission rates as high as 1600 to 2000 per 1000 troops per year were observed. In rifle companies in the North African Theatre during the Tunisian, Sicilian, and Italian campaigns, neuropsychiatric casualty rates reached 1200 to 1500 per 1000 strength per year. The First Army in Europe had a division that suffered 944 neuropsychiatric casualties out of 1100 total medical admissions, while in eight divisions there was an average of 200 neuropsychiatric casualties out of 482 total medical admissions per year during the first two months after D-Day in France. During June through July 1944, the First Army reported 10,081 neuropsychiatric admissions and 55,517 battle casualties, half of whom were classified as serious.

The ratio of admissions for neuropsychiatric casualties to wounded in action casualties has been on the order of one to three for the duration of the entire conflict. However, in high intensity conflicts of short duration, this 1 to 3 ratio has been changed to more stress reaction casualties than WIAs. Forissier and Darmandieu (1976) report in the 1973 Arab-Israeli War that 900 of the initial 1500 evacuated Israeli casualties were not physically injured. The nature of modern warfare has changed: the number of pulses of battle has moved from the four to five pulses per day of World War II to the sustained, mobile, continuous engagements with 10 to 12 pulses per day over extended durations. Continuous operations will accelerate the onset of combat stress reactions.

Troops do not have to be in the combat zone to suffer stress reactions. As far back as World War I, many of the symptoms quoted as "war symptoms" were observed in the home cantonments during training. The change from civilian to military life in itself may be sufficiently stressful for some individuals. Troops that have just recently entered the service are generally greater risks for having stress reactions than soldiers with two or more years of military service.

In the Southwest Pacific Theatre in World War II, there were more admissions for neuropsychiatric casualties per 1000 men strength per year than battle injuries. The factors involved in the Southwest Pacific area were: the tropical climate, monotony of the jungle and environment, and lack of adequate rotational policy. During the Korean War, relatively few psychiatric situational disorders were admitted for troops in the combat support zone. However, troops assigned to the rear echelon group in Korea were somewhat less effective in their performance than combat support zone or front-line combat troops. The less effective performance was attributed to boredom, frustrations of monotonous living, lack of visible importance in overall effort. Similarly in Vietnam, support units had trouble perceiving the usefulness of their work. Support troops may be more susceptible to stress reactions.

Stress reactions in or out of the combat situation represent adaptive mechanisms for coping with what might be termed abnormal circumstances. If troops expect that breaking down in combat is indicative of being mentally ill, the manifestations of that suggested expectation may take the form of stereotyped, abnormal behaviors. Psychiatric labels for stress reactions may lead both the combat group and the treatment providers to expect bizarre types of behavior. However, if the combat group refuses to allow maladaptive responses among its individual members, non-effective behaviors in response to stress can be minimized.

The labels used for neuropsychiatric casualties have a powerful suggestive influence upon the health care providers as well. If combat neuropsychiatric casualties are seen as suffering from "shell shock," "psychoneuroses," "war neuroses," or "concussion," it will be difficult to foster the expectation of successfully returning the soldier to duty. During World War II, new terminology was introduced to suggest less pessimistic labels such as "combat exhaustion" and "combat fatigue." These labels suggested a more rapid recovery was possible from a stress response.

Through trial and error, it was discovered that troops who had suffered stress reactions in combat were most effectively treated as near to the front lines as possible. These troops were reassured and provided the expectations of returning as soon as possible to join their unit at the front. The principles of (1) proximity, (2) immediacy, and (3) expectation were developed during World War I, forgotten between the wars, re-learned in World War II, partially recalled in Korea, and forgotten but re-learned in the 1973 Israeli war.

Knowing the treatment principles is not sufficient. By being aware of why the principles apply, the health care provider can more effectively help to conserve the fighting strength. Since individual soldiers differ in their vulnerability to stress reactions, a soldier should be considered as an integral part of his group unit. Influence from group cohesion, unity, leadership, previous battle training, and experience help to preserve the individual soldier's effectiveness as part of his group. In stressful situations, the soldier compares himself with similar others who are facing comparable stresses. The group members will strive likewise to perform effectively. The group dynamics influence the individual's behavior (Festinger social comparison theory, 1954).

Schacter (1962) recognized the importance of cognitive appraisal of physiological responses to emotional situations. If a person is aroused physiologically, what his mind tells him about labeling the situation can influence his emotional response. If the soldier expects to rejoin his group, he will try. Forward therapy can be successful because the casualty still has strong emotional ties to his unit and group identification. The phenomenon of expectancy was first noted in World War I when it was observed that recent psychiatric casualties were quite suggestible. With firm direction, the suggestible soldier could be influenced toward improvement and return to duty. Medical treatment personnel are critical in providing a calm acceptance of the soldier's symptoms. Reassurance that the response is a temporary breakdown in response to the stresses of combat, together with rest and shelter, can produce rapid improvement.

As mentioned, the treatment personnel are critical in fostering and maintaining the expectation that the soldier who has had a stress reaction will recover and return to duty. The treatment personnel must believe the concept and perceive themselves as part of the unit. If the soldier is allowed to be evacuated back, his identification with his unit declines as does his confidence in his own abilities. The medical evacuation of neuropsychiatric casualties leads to the fixation of symptoms and helplessness. The further back an individual is removed through evacuation, the smaller the chance for his return to duty and/or to effective functioning.

In combat, there may be many stress reactions which are taken care of by the individuals in a group unit. In units that have strong group identification, cohesiveness, leadership, and morale--there are few reports of stress reactions. Elite units sustain the individual members' abilities to cope with stress. There are no statistics available for the actual number of stress reactions in combat units.

Emphasis is treatment as far forward as possible. Separate channels are to be maintained for the movement of combat stress reactions to be distinguished from WIAs. Upon arrival at a Battalion Aid Station or Brigade Area/Clearing Station, it will be impossible to accurately diagnose soldiers. Careful observation of the soldiers will allow for determining which management procedures can be employed. Rest, food, relative safety, positive expectation of return to effective functioning in combat, opportunities for ventilation about combat experiences, directive counselling focusing on symptom alleviation, and reassurance should be employed. The soldier, his unit, his peers, the treatment personnel must all expect and assure the CSR he will return to duty.

The most effective means of prevention is to create conditions of high unit morale and cohesiveness. This can be accomplished through effective training, leadership, realistic expectations, and frequent interactions with treatment personnel. Individuals and units must know what is expected of them; individuals must be aware of what types of stresses they may experience and be prepared through continual training to develop emotional-sustaining behaviors to cope with these situations. Mental health officers and enlisted personnel must actively take part in the training. Mental health personnel must properly educate commanders, NCOs, and their units of the possibility of stress reaction symptoms as normal responses to abnormal situations. The mental health personnel must be perceived as part of the unit and be actively involved in the combat division training.

Unit members must be supportive of one another in stress situations, whether in realistic training or in combat. Well-practiced behaviors (immediate action drills) must be automatic responses in stress situations. No effective behaviors may cause needless loss of life. Preventive education and mental health consultation can make the difference in making available manpower replacements.

Units must expect to take care of their own members in stress situations. Individual combat stress reactions that cannot be handled at the local unit level by the NCOs or unit commanders should be sent to the Battalion Aid Station for rest, observation, and limited psycho-therapy. A mental health officer at the Division Clearing Station should be employed as a consultant to the more forward treatment areas.

Currently, there are several programs under way to train medical corpsmen (91Bs, 91Cs, 91Gs), mental health professionals, NCOs, and officers to recognize that combat stress reactions may affect all soldiers, particularly those in positions of responsibility. Workshops like Project COPE (Combat Operations and Psychiatric Effectiveness) and the Users' Workshop on Combat Stress at the Academy of Health Sciences are focused on providing useful training packages through experiential learning. Future workshops will be oriented toward bringing together commanders, division surgeons, and mental health professionals to set goals to train to recognize and prevent combat stress reactions.

References available on request.

Effect of Individual and Organizational Variables
on Perceived Job Stress

William H. Hendrix and Nestor K. Ovalle, 2d
Department of Organizational Sciences
Air Force Institute of Technology
Wright-Patterson AFB OH

Raymond G. Troxler
Brooks Medical Research
Institute
Brooks AFB TX

Charles W. Staton
Headquarters Air Defense Tactical Air Command
Langley AFB VA

Samuel P. Fye
USAF Hospital
Reese AFB TX

Abstract

Development and validation of the Stress Assessment Package is discussed. Factors and background information items were regressed against individual's perceived stress factor. Results indicated a series of individual and organizational factors, and background items were related to stress. These provide a basis for individuals and organizations to better manage undesirable stress.

Job stress has become a topic of increasing interest to organizations. Stress can have severe effects on organizational effectiveness, both short term and long term. In the short term if an individual's stress is continually increased a point will be reached where performance decreases as stress level increases. Along with this decreased performance for the individual is the likelihood that group performance will also decrease due to the individual's reaction. The individual may not have time to interact properly with group members. This behavior may take the form of developing a short temper or hoarding information needed for task accomplishment. In addition, the individual may develop physical problems which decrease his or her effectiveness on the job, or if severe enough may require hospitalization. The problems of ulcers, high blood pressure, allergies, and coronary heart disease are believed to be in part precipitated by stress.

Various physiological changes occur when one is exposed to a stressful environment. Two blood components affected by stress are cholesterol and cortisol (an adrenal hormone). Friedman and Carroll (1957) examined tax accountants to determine the effects that heavy workload, high level of responsibility, time pressure, conflict, and job-role ambiguity had on cholesterol level. Their results indicated that there was a marked increase in the blood cholesterol level as the tax-filing deadline approached. After the deadline passed, the cholesterol decreased; returning to normal within two months.

HDL cholesterol, on the other hand, has been indicated as a coronary heart disease reducing factor (Kritchevsky, Paoletti, and Holms, 1978). That is, as HDL cholesterol increases, there is a decreasing probability of developing coronary heart disease.

Cortisol is an adrenal hormone which is secreted into the blood stream. A series of studies (Brown, Schalch, and Reichlin, 1971; Kopin, 1976; Rubin, Rache, Clark, and Arthur, 1970) have indicated that as stress increases there is a resulting increase in the blood cortisol level. In addition, there is some evidence that increased cortisol levels result in increased total cholesterol levels (Troxler, 1980). This relationship suggests that stress may be a factor in the development of coronary heart disease.

This study is a part of a large-scale stress research program being conducted by the Air Force Institute of Technology and the School of Aerospace Medicine. The program includes the development of a Stress Assessment Package to measure personality, organizational and background factors which are related to stress and coronary heart disease. These survey factors are to be related to physiological measures of stress (cortisol) and coronary heart disease potential (total cholesterol ÷ HDL cholesterol).

This study is limited to investigating which factors contained within the Stress Assessment Package were related to perceived stress.

Method

Subjects

A sample of 363 individuals completed the Stress Assessment Package. With the exception of approximately 40 employees of a civilian hospital, all individuals were Air Force civilian and military employees at either Eglin Air Force Base, Kelly Air Force Base or Wright-Patterson Air Force Base. Approximately 85% were males and 15% were females. Participation was on a voluntary basis and anonymity was insured by each subject selecting a number which served as their personal identifier known only to them.

Survey Instrument

The Stress Assessment Package was used for data collection. This survey consisted of 139 items of which 110 were primarily 7-point Likert attitudinal scales and 29 were background information items. The Likert attitudinal items were designed to measure organizational variables (e.g., organizational climate, job enrichment, autonomy, role conflict, and goal setting) and personality variables (e.g., Type A Behavior and Locus of Control). The background information items were used to collect data such as sex category, race, and for personal history items such as smoking and jogging experience.

Procedure

The Stress Assessment Package was administered to volunteers en masse at each administration site. After completing the survey, individuals computed their indices on a series of factors such as assertiveness and Locus of Control. An explanation of these factors and how each was related to stress was provided. Individuals desiring to learn of their cholesterol and cortisol levels had their blood drawn. Almost all individuals completing the survey also had their blood drawn (approximately 97%). The attitudinal items were factors analyzed with 18 orthogonal factors extracted (Table 1). In turn,

a perceived job-stress scale was used as the dependent variable for assessing which personality and organizational factors were related to perceived job stress. In addition, background information items and attitudinal items which did not load heavily on the 18 factors were regressed against the perceived stress scale.

Insert Table 1 about here

Results and Discussion

The question under investigation: "What organizational and individual factors are predictive of perceived job stress?", was answered by the stepwise regression analysis. The first used individual factor scores for the 18 factors extracted as inputs to a regression analysis. The results are provided in Table 2. These results indicate that individuals experience more stress if they are a Type A - External Locus of Control individual, who works in an organization that is characterized by high inter-group conflict and role conflict, poor or low job satisfaction and organizational climate, and if they were actively involved in goal-setting activities.

Insert Table 2 and 3 about here

The results of the second regression analysis which used the individual attitudinal and background information items is provided in Table 3. The results indicate increased stress occurs when there are inadequate materials to accomplish one's job; quantity of output is emphasized over quality; one has increased responsibility; one is overweight; and for civilian employees not employed by the federal government, stress increased with an increase in organizational level. Stress also decreased for those who jogged, had high-job security, performed acceptable tasks on the job, had adequate job equipment, and good organizational communications.

Together these analyses provide a basis for individuals to assess their standing on factors related to stress. The Stress Assessment Package provides them with indicators for use in modifying their behavior or job environment to decrease undesirable stress. For the organization, its relative standing on the stress-related factors/items provides a basis for organizational modifications to reduce undesirable stress.

Continuing research should provide an even firmer basis for stress evaluation. This research will link the Stress Assessment Package factors to the physiological indicators of stress (i.e., cortisol) and potential for coronary heart disease (i.e., total cholesterol ÷ HDL cholesterol). Hopefully, these efforts will provide individuals and organizations with better insight into one component which effects the quality of worklife and organization effectiveness: Stress!

References

- Brown, G. M., Schalch, D. S., & Reichlin, S. Patterns of growth hormone and cortisol responses to psychological stress in the squirrel monkey. Endocrinology, 1971, 88, 956-963.
- Friedman, M., & Carroll, V. Changes in the serum cholesterol and blood clotting time of men subject to cyclic variation of occupational stress. Circulation, 1957, 17, 852-861.
- Konlin, I. J. Catecholamines, adrenal hormones, and stress. Hospital Practice, March 1976, 49-55.
- Kritchevsky, D., Paoletti, R. & Holms, W. Drugs, lipid metabolism, and atherosclerosis, New York: Plenum Press, 1978.
- Rubin, R. T., Rache, R. H., Clark, B. R., & Arthur, R. J. Serum uric acid, cholesterol, and cortisol levels: Interrelationship in normal men under stress. Archives of internal medicine, 1970, 125, 815-819.

TABLE 1. Orthogonal Factors Extracted

Factor No.	Factor	Cronbach Alpha
1	Job Enhancement	.90
2	Supervision	.92
3	Assertiveness	.82
4	Productivity	.91
5	Internal/External Locus of Control	.73
6	Organizational Climate	.84
7	Need for Enrichment	.83
8	Type A/Type B Behavior Pattern	.66
9	Time Adequacy	.71
10	Rules and Regulations	.82
11	Goal Clarity	.77
12	Job Autonomy	.90
13	Job Importance	.90
14	Goal Setting	.90
15	Availability of Time to Plan	.76
16	Role Conflict	.77
17	Work Information	.84
18	Intergroup Conflict	.71

TABLE 2. Regression Analysis Using Factor Scores

Factor	Factors' Relationship to Stress	R	R ²
Type A Behavior	+	0.199	0.040
Inter-group Conflict	+	0.274	0.075
Locus of Control (External)	+	0.322	0.104
Role Conflict	+	0.353	0.125
Job Satisfaction	-	0.376	0.142
Goal Setting	+	0.399	0.159
Organizational Climate	-	0.417	0.174

TABLE 3. Regression Analysis Using Individual Items

Factor	Items' Relationship to Stress	R	R ²
Inadequate Material	+	0.445	0.198
Grade Level (Wage Grade)	-	0.481	0.231
Jogging	-	0.509	0.259
Job Security	-	0.535	0.286
Quantity of Output	+	0.553	0.305
Group Meetings	+	0.569	0.324
Responsibility for your Own Work	+	0.580	0.336
Officer Grade Level	-	0.596	0.355
Weight	+	0.609	0.371
Acceptable Tasks	-	0.617	0.381
Adequate Equipment	-	0.625	0.390
Communication	-	0.632	0.399
Civilian Employee (non-DOD)	+	0.638	0.407

A Pilot Evaluation of Stress Management Techniques for Collegiate Rifle Shooters

Kenneth C. Coon & Richard L. Hughes
U.S. Air Force Academy

Abstract

Members of a nationally ranked intercollegiate rifle team were assigned to either of two different stress management treatment groups. One group received six sessions of relaxation training and the other received six sessions based upon a cognitive model of stress management. At posttesting neither group had improved shooting score in match-like conditions nor had shooters' subjective tension in competition decreased. Methodological problems which affected these results are discussed.

Everyone from the Sunday golfer agonizing over a two foot putt to a professional basketball player getting set to take a foul shot in a critical game has probably felt the stress of competition. The pressure of athletic competition can be highly intense and has led to a profitable industry in sports psychology and stress management. However, relatively little systematic research has compared the effectiveness of specific approaches to managing stress in athletic situations.

In general, the literature contains uncontrolled and impressionistic evaluations of stress management programs, case studies of successful interventions with individual athletes (Nideffer & Deckner, 1970; Garver, 1977), and articles in special interest publications (e.g. Golf Magazine) offering advice or formulas for "how-to" manage stress (Harrison & Mitchell, 1978; Peterson, 1977).

Stress management techniques can be roughly categorized into two broad approaches which, following a distinction elsewhere (Liebert & Morris, 1967), may be termed a physiological approach and a cognitive approach. Such techniques as progressive relaxation, hypnosis, biofeedback, transcendental meditation, autogenic training and systematic desensitization fall in the former grouping in that stress is conceptualized in terms of physiological overarousal and remediated by reducing arousal level. The cognitive approach, on the other hand, attributes stress to dysfunctional attentional processes or self-defeating attitudes and remediation is directed toward developing constructive ways of thinking in stressful situations. This approach seems well-suited to stress management in athletic competition, though it has been primarily applied and evaluated in other sorts of stress situations (Holroyd, 1976; Meichenbaum, 1972).

Another cognitive approach is mental practice, which is the covert rehearsal of specific athletic behaviors (Suinn, 1977). Mental practice has been shown to improve performance on motor skills (Richardson, 1972) and is used by many coaches and athletes. It often includes vivid visualization of one's performance with attending covert kinesthetic and motor involvement.

Virtually all of these procedures have been systematically evaluated in other contexts, but they have not been compared as treatments of competition stress in athletic situations. Such a comparison seems advisable since the history of stress management research in the widely researched area of test anxiety records distinct effects for relaxation and cognitive treatments.

Relaxation-based treatments were long the treatment of choice for remedying test anxiety, but research gradually indicated that reductions in test anxiety were not matched by comparable improvements in academic performance (Spielberger, Anton & Bedell, 1976). Cognitive modification, on the other hand, has been a more effective technique for improving performance. Therefore, it seemed advisable to evaluate the relative effectiveness of relaxation and cognitive approaches in an athletic context.

Another reason why the application of stress management in sports warrants systematic evaluation is because "sports" really comprise a heterogeneous grouping of activities which vary, among other ways, in terms of desirable arousal levels for optimal performance (Oxendine, 1970). Some sports (e.g. weight lifting) are suited for high arousal levels while others (e.g. archery, putting in golf, compulsories in figure skating) are suited for low arousal levels. Given such differences, it seems reasonable that stress management procedures should be tailored for the characteristics of each sport.

The present study involved the comparison of two approaches to stress management with a nationally ranked intercollegiate rifle team. Rifle would seem to be an ideal sport for evaluating such an intervention. First, individual performance can be objectively measured. Second, shooting is a mental sport; after a shooter gains a basic level of physical proficiency in shooting, concentration and calmness during shooting become critical to performance. Since it is literally a sport where fractions of an inch make a big difference in score, the tension created by one bad shot can disrupt the rest of one's match.

METHOD

Subjects

The subjects in this study were 18 members of the intercollegiate varsity rifle team at the U.S. Air Force Academy. Each team member was randomly assigned to one of two treatment groups which were balanced for shooting ability.

Stress Management Treatments

The stress management techniques were presented by two masters-level professional counselors who each had several years of counseling experience. Each counselor conducted one relaxation group and one cognitive modification group.

The relaxation training sessions were presented primarily via a series of recorded audio tapes. Deep relaxation techniques were practiced by team members as they listened to the tapes. The tapes included breathing exercises, progressive relaxation exercises for contrasting muscle tension and relaxation, and techniques for smooth muscle relaxation. The counselors described the applicability of relaxation techniques to stressful situations and answered

questions, but the bulk of the relaxation treatment sessions consisted of the standardized cassette tapes.

The cognitive modification sessions involved presentation of an information-processing model of stress wherein dysfunctional cognitions - catastrophizing, negativistic, self-defeating - are conceptualized as central in causing stress. The treatment emphasized the value of positive self-talk and provided specific techniques for controlling negative self-talk. Additionally, these sessions included mental practice of certain key shooting skills (e.g. coordinating breathing and trigger pressure) in imagined competition, and the use of self-instructions as cues for achieving optimal arousal levels.

Treatment consisted of six 50-minute sessions presented weekly for six consecutive weeks to each group.

Dependent Variables

Three types of measures were used to assess the success of the stress management interventions. These measures were shooting scores, subjective tension during match-like conditions, and shooters' critiques of the program's value.

RESULTS

In general, there were no significant effects observed for either the relaxation or the cognitive stress management interventions on any dependent variable. In fact, there was a slight (nonsignificant) decline in shooting scores between the pre and posttesting and likewise a slight (nonsignificant) increase in subjective tension at posttesting. Although new shooters (freshmen) improved more in shooting score than did experienced shooters, this likely represented the effect of intervening practice and not an effect of the intervention itself. The results of the program critiques were somewhat more positive. On a 7-point scale assessing the program's value in improving stress management in rifle competition, the mean response was slightly positive (\bar{X} 4.4). But most telling of all was the variation in the evaluations which ranged across nearly the whole scale.

DISCUSSION

The shooters' critiques may be summarized as indicating that everything was helpful for somebody, but nothing was helpful for everybody. With such considerable individual variation in response to the interventions, and without any overall significant effects, it may be most appropriate here to relate our general observations and recommendations for any similar planned interventions.

First, our objective to conduct a methodologically sound evaluation of the two treatments was only partially met. In order to maximize the size of our treatment groups we opted to include virtually all team members in the study. However, that involved including some who by any measure did not need stress management.

Second, the impact of the program could have been enhanced by more frequent sessions than once per week as well as by providing structured materials (outlines, homework assignments, etc.) for athletes to keep.

Although a few shooters described aspects of the mental practice as helpful, it may be that rifle shooting is simply too static a sport to benefit much from mental practice. Mental practice may be more appropriate for skills involving larger bodily movements.

As in other areas of stress research (e.g. test anxiety), it proved difficult to simulate the level of stress often felt in an actual match. Stress management skills are most likely to be applied in competition if they have been practiced earlier in maximally similar conditions, but the difficulty in simulating in practice the stress induced by (e.g.) a bad opening shot in an actual match makes such generalization difficult.

For a similar reason, we must question the meaning of our own tension self-reports, which were obtained at the beginning of the practice matches. It seemed to be the case that shooters were generally unstressed at the outset of these matches but sometimes became anxious upon making a bad shot. In retrospect, we seem to have mistimed our questions, and the reports do not reflect subjective tension at the critical times.

In sum, despite attempts to carefully evaluate two different approaches to stress management in athletic competition, several internal weaknesses of design limited the effects which could be observed. However, these problems in no way contradict the notion that stress management evaluation research in athletic situations is pragmatically and theoretically necessary - they only reaffirm the fact that it's difficult!

REFERENCES

- Garver, R.B. The enhancement of human performance with hypnosis through neuromotor facilitation and control of arousal level. The American Journal of Clinical Hypnosis, 1977, 19(3), 177-181.
- Harrison, W. and Mitchell, M. Target your thoughts. Golf Magazine, 1978, 74-77.
- Holroyd, K.A. Cognition and desensitization in the group treatment of test anxiety. Journal of Consulting and Clinical Psychology, 1976, 44(6), 991-1001.
- Liebert, R.M. and Morris, L.W. Cognitive and emotional components of test anxiety: A distinction and some initial data. Psychological Reports, 1967, 20, 975-978.
- Meichenbaum, D.H. Cognitive modification of test anxious college students. Journal of Consulting and Clinical Psychology, 1972, 39(3), 370-380.
- Nideffer, R.M. and Deckner, C.W. A case study of improved athletic performance following use of relaxation procedures. Perceptual and Motor Skills, 1970, 30, 821-822.
- Oxendine, J.B. Emotional arousal and motor performance. Quest, 1970, 13, 23-30.
- Peterson, B.A. Overcoming shooting slumps. Bow and Arrow, 1977, 15, 47-49.
- Richardson, A. Voluntary control of the memory image. In P.S. Sheehan (Ed.), The function and nature of imagery. New York: Academic Press, 1972.
- Spielberger, C.D., Anton, W.D., and Bedell, J. The nature and treatment of test anxiety. In M. Zuckerman & C.D. Spielberger (Eds.), Emotion and Anxiety: new concepts, methods and applications. New York: Wiley, 1976.
- Suinn, R.M. Body thinking: Psychology for Olympic champs. Swimming Technique, 1977, 14, 8-21.

The Desirability of Alternatives:
A Promising Construct for Understanding Turnover Decisions

Thomas W. Watson
Air Force Human Resources Laboratory, Brooks AFB TX

Victor H. Appel
University of Texas at Austin, Austin TX

Abstract

The construct, desirability of the alternatives, is presented as a promising perspective for predicting and explaining turnover decisions. The nature of this construct is discussed, and theoretical and empirical support for its importance to turnover decisions is presented. Implications for turnover theory and research are also discussed.

Employee turnover has been historically a topic of concern to personnel interested in organizational behavior. For at least a half century, efforts have been made to explain the basis of turnover within civilian and military contexts. However, as late as 1978, the most sophisticated studies were accounting for only 25% of the turnover variance. In this paper a promising construct, the desirability of alternatives, is introduced to provide a framework for the study of turnover decisions. This construct has already allowed investigators to account for a far greater proportion of the turnover variance than was able to be accounted for in earlier studies. For example, using a parsimonious model containing a desirability-of-alternatives variable which he referred to as environmental pull, Bluedorn (1979) was able to account for 65% of the turnover-intent variance in a study involving Army officers. Also, as the social-reinforcement exchange theory of Thibaut and Kelley (1959) suggests, such a construct can be applied to other decision-making contexts. These theorists applied a similar construct to decisions concerning the maintenance or termination of dyadic relationships. However, despite the applicability of the desirability-of-alternatives construct to multiple decision-making situations, this paper will focus primarily on turnover decisions.

The problem of turnover among military personnel is well known. Since involuntary conscription was terminated in the early 1970's, recruitment and retention of military personnel have become increasingly critical issues. In fact, the Air Force and its sister services consider the maintenance of an adequate military force to be of primary concern during the 1980's. However, recent advances in theory and research indicate that the utility of the desirability-of-alternatives construct may even extend to non-military contexts. Thus, this paper should be relevant to diverse groups of researchers and managers interested in turnover, regardless of their affiliations.

Nature of the Desirability-of-Alternatives Construct

As stressed by Mobley, Griffeth, Hand, and Meglino (1979), in their comprehensive review of the turnover literature, a growing consensus is emerging among investigators that the explanation and prediction of turnover should be viewed from the perspective of the desirability of alternatives. Such an approach acknowledges that reenlistment (or the decision to remain in or to leave any context) should be construed as a decision-making situation (see Mitchell & Beach, 1976). In such decision-making situations, an individual weighs his or her future prospects within an existing setting against comparable opportunities perceived to be available elsewhere. Such a perspective gives implicit recognition to the fact that the merits of a given setting cannot be meaningfully assessed apart from simultaneous assessment of the merits of available options. Thus, consideration of the desirability-of-alternatives construct stresses the need to take a broad systems perspective in which focus is shifted from assessment of absolute satisfaction in a single context to assessment of relative satisfaction in at least two contexts.

Historically, satisfaction has been a key variable used to predict and explain turnover (see Brayfield & Crockett, 1955; Porter & Steers, 1973; and Tuttle & Hazel, 1974, for reviews of this literature). However, the correlation between satisfaction and turnover has typically been found to be moderate, and, as Mobley et al. (1979) have concluded, studies using satisfaction as the sole predictor variable tend to account for not more than 16% of turnover variance. The importance of the desirability-of-alternatives construct, and its emphasis on relative rather than absolute satisfaction, can best be demonstrated by the failure of the satisfaction hypothesis to predict turnover behavior easily accounted for from a desirability-of-alternatives perspective. By considering the desirability of an obtainable alternative, the paradoxical finding that some satisfied persons leave a setting, rather than stay as expected, while some dissatisfied persons remain, can be explained.

Theoretical and Empirical Support for the Desirability of Alternatives

Theoretical Support

General theoretical support for the importance of the desirability of alternatives as a perspective from which to view decision making has been accumulating for some time. The theory of dyadic interaction proposed by Thibaut and Kelley (1959) more than 2 decades ago has already been mentioned. Although this theory emphasizes the importance of perceived alternatives in the maintenance or termination of person-to-person relationships, the present authors posit that such a construct can be applied, by extension, to the relationship between an individual and an organization.

More recently, with specific reference to organizational behavior, Smith, Kendall, and Hulin (1969) foreshadowed the need to consider the perception of alternatives in the understanding and prediction of behavior. It is interesting to note that such a multiple-systems perspective was advocated by a group of highly respected investigators after they had spent years studying satisfaction without taking such a perspective into account.

Despite the recommendation concerning alternatives of Smith and her associates more than a decade ago, it is only recently that theorists interested in turnover have emphasized the importance of the desirability-of-alternatives construct. However, Price (1977), on the basis of an extensive literature review, recently proposed a parsimonious model of turnover which stressed the importance of opportunity, his term for the desirability-of-alternatives construct discussed in this paper. He proposed that if satisfaction with an existing work setting is low, and opportunity is high, the likelihood of turnover will be great. Similarly, in their recent review of the literature, Mobley and his associates (1979) concluded that the evaluation of perceived alternatives appears to be a crucial variable in the turnover process. More recently, Greenhalgh (1980) also contributed theoretical support for the desirability-of-alternatives construct. In proposing a developmental process model emphasizing changes in the factors influencing turnover decisions over time, he contended that the decision to leave an organization typically will be implemented only if positive opportunities exist in the environment external to the organization.

Empirical Support

As the brief review of theoretical support for the desirability-of-alternatives construct provided above indicates, the possible importance of such a variable is not a new phenomenon. However, investigators interested in turnover only recently (circa 1976) began to incorporate this variable in their research. Thus, empirical evidence in support of the concept has only begun to accumulate. Such evidence is selectively reviewed below.

Schneider (1976) investigated the relative attractiveness of existing Navy jobs versus civilian alternative positions and was one of the first investigators interested in turnover to employ a desirability-of-alternatives approach. He referred to the desirability-of-alternatives construct as the "greener-grass phenomenon."

Schneider emphasized the comparison of various attributes of Navy jobs versus alternative civilian positions almost to the exclusion of other variables. In contrast, Bluedorn (1979), who was interested in Army officer turnover, took a different approach. He included environmental pull (his term for the desirability-of-alternatives construct) as one of five predictors of turnover intent. Using such a parsimonious model, he was able to account for 65% of the turnover-intent variance, a sizable increase over the maximum of about 25% accounted for in earlier bivariate studies. Also, Bluedorn found environmental pull to rank first among the variables included in his model in terms of effect on turnover-intent decisions.

The theoretical contribution of Price (1977), who proposed a theoretical model similar to Bluedorn's, has already been mentioned. Price and Mueller (1981) extended the model developed by Price (1977) and included opportunity (their term for the desirability-of-alternatives construct) as one of 11 determinants of turnover. It is noteworthy that opportunity was the second most important determinant, after intent to stay. Price and Mueller further stressed the importance of opportunity by pointing out that, along with pay, it is the most widely supported turnover determinant in the literature.

However, despite the support for pay as a determinant of turnover, they found that opportunity was approximately four times as important as pay in predicting turnover. Such a finding further emphasizes the relevance of the desirability of alternatives in the study of turnover.

Implications for Theory and Research

The premise advanced by Mobley and his associates that the evaluation of alternatives is apparently a crucial variable in the turnover process is well supported in both the recent theoretical and empirical literature. It can be concluded that models of turnover should include such a construct and that more emphasis should be placed on the desirability-of-alternatives construct in future research. Also, as studies by Bluedorn (1979) and Price and Mueller (1981) indicate, multivariate rather than bivariate methods should be used to account for greater proportions of the turnover variance. This point was also stressed by Mobley et al. (1979) in their recent literature review.

REFERENCES

- Bluedorn, A. C. Structure, environment, and satisfaction: Toward a causal model of turnover from organizations. Journal of Political and Military Sociology, 1979, 7, 181-207.
- Brayfield, A. H., & Crockett, W. H. Employee attitudes and employee performance. Psychological Bulletin, 1955, 52, 396-424.
- Greenhalgh, L. A process model of organizational turnover: The relationship with job security as a case in point. Academy of Management Review, 1980, 5, 299-303.
- Mitchell, T. R., & Beach, L. R. A review of occupational preference and choice research using expectancy theory and decision theory. Journal of Occupational Psychology, 1976, 49, 231-248.
- Mobley, W. H., Griffeth, R. W., Hand, H. H., & Meglino, B. M. Review and conceptual analysis of the employee turnover process. Psychological Bulletin, 1979, 86, 493-522.
- Porter, L. W., & Steers, R. M. Organizational, work, and personal factors in employee turnover and absenteeism. Psychological Bulletin, 1973, 80, 151-176.
- Price, J. L. The study of turnover. Ames: Iowa State University Press, 1977.
- Price, J. L., & Mueller, C. W. A causal model of turnover for nurses. Academy of Management Journal, 1981, 24, 543-565.
- Schneider, J. The "greener grass" phenomenon: Differential effects of a work context alternative on organizational participation and withdrawal intentions. Organizational Behavior and Human Performance, 1976, 16, 308-333.
- Smith, P. C., Kendall, L. M., & Hulin, C. L. The measurement of satisfaction in work and retirement. Chicago: Rand-McNally, 1969.
- Thibaut, J. W., & Kelley, H. H. The social psychology of groups. New York: Wiley, 1959.
- Tuttle, T. C., & Hazel, J. T. Review and implications of job satisfaction and work motivation theories for Air Force research (AFHRL-TR-73-56, AD-782 443). Lackland Air Force Base, TX: Air Force Human Resources Laboratory, Occupational Research Division, January 1974.

EXAMINING THE TURNOVER PROCESSES
WITHIN THE FRAMEWORK OF AN "AFFECT
MINIMIZATION" PARADIGM

Gerald D. McCright
University of Texas, Austin

Abstract

An affect-minimization model of the turnover process is proposed. The conceptualization incorporates cognitions of the relevant resident and alternative environmental factors with potential behaviors as a means of inducing an affect continuum. It is thence hypothesized that behavioral intentions correspond to the minima of the induced affect continuum. An affect "function", which is mathematically equivalent to Thom's (1975) butterfly catastrophe potential function, is introduced and tested with a small sample (n=60).

Mobley (1982) emphasizes that turnover is a unique event in time, which is preceded by a host of turnover antecedents, and is followed by often-ignored turnover consequences. He casts the problem within a total psychological context, asserting that turnover involves cognitions and affects as well as behaviors. Within that broad context, Mobley contends that greater attention must be accorded the "process" in turnover-process research, noting that static models of turnover necessarily fail to capture its dynamic character. He characterizes turnover as a time-connected sequence of changing factors and relationships among factors conducing to the turnover event which can properly be examined only by longitudinal means. Implicit in Mobley's argument is that the time-associated changes will somehow be related in an orderly and systematic fashion. That is, turnover is perceived to be not only dynamic, but it is perceived to function in the manner of a dynamic system. And, indeed, the system property that is being invoked is stability. Hence, one may legitimately ask, what, if anything, lawfully "drives" the system and maintains the (implied) stability?

Within dynamic systems theory, one of the crucial concepts is stable equilibrium states of the system which, in general, are associated with the minimization of some fundamental property of the system, e.g., total pressure or potential energy (cf. Zeeman, 1977; Tanaka, 1981). If such a concept is invoked within the turnover system it resurrects a model which was initially proposed four decades ago. Every psychologist is familiar with Dollard and Miller's "approach-avoidance conflict" conceptualization (Miller, 1944; Dollard & Miller, 1950); a simple, yet classic, example of a dynamic, psychological-behavioral system with an induced equilibrium state. It is apparent upon examining their model that conceptually it may be perceived as locally minimizing total affect, i.e., fear plus desire, at some distance from the defined goal. Movement either toward or away from the goal subsequently generates a restoring "force" to maintain the stable

distance from the goal. Despite the severe limitations of the Dollard & Miller model, might we not be wise to make use of their concept of affect minimization as the means of inducing stable equilibrium in an essentially three-state dynamic system of turn-over decision making. Define state I as a, "preference for the resident environment"; state II as a, "preference for an alternative environment"; state III as, "avoidance of resident and alternative environments." Designate the states as E_r , E_a , and E_o , respectively. Figure 1 is a schematic representation of the foregoing conceptualizations.

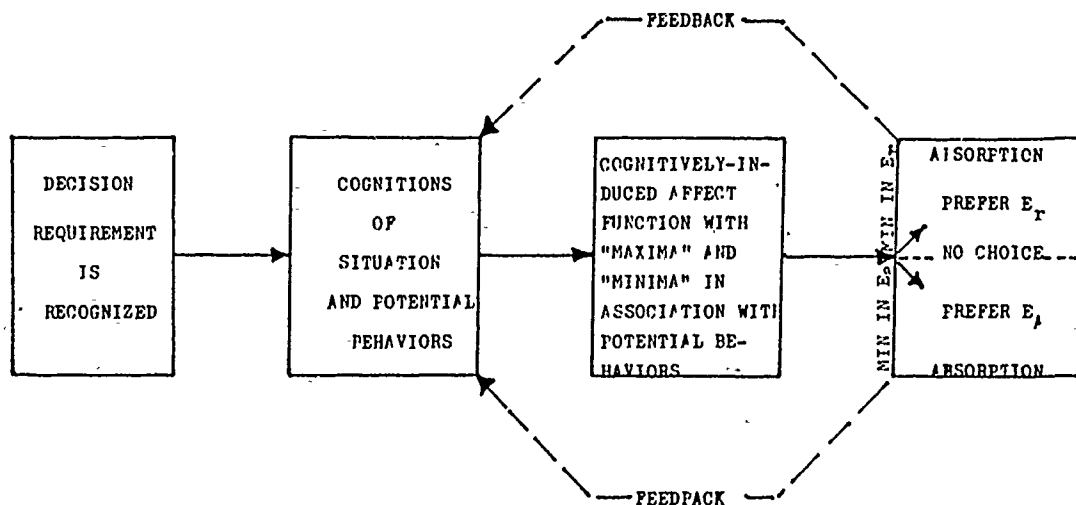


Fig. 1. Conceptual model of Cognition-Affect-Behavior dynamic system

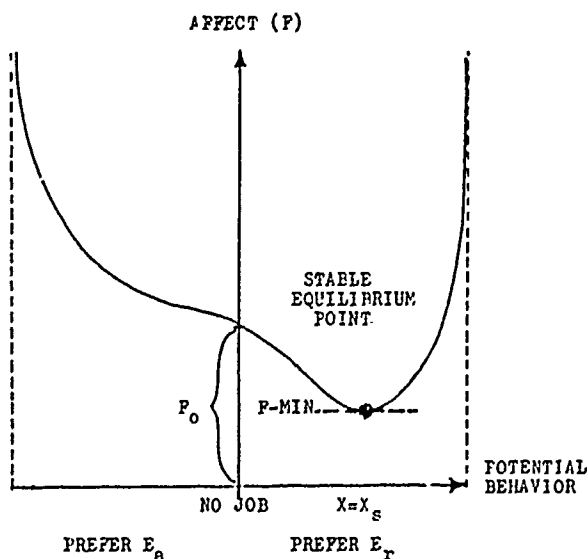


Fig. 2. Schematic representation of hypothesized cognitively-induced "affect function" (P). In this example, the affect function exhibits a minimum value within the "Prefer E_r " behavior state.

In fig. 1 the system is initialized by recognition of the decision requirement. Thence, assessment of the relevant situational factors maps onto the (cognized) potential-behavior continuum. This mapping is posited to induce an "affect continuum" isomorphic to the behavioral continuum. This result is illustrated in fig. 2. We then invoke a variant of the Dollard & Miller behavior algorithm, viz., move to and remain at the nearest affect minimum. The "affect-gradient-induced" preference then feeds back to the cognition block and the process iterates.

There are four distinguishable geometric forms of the posited affect function:

- I: An affect minimum within E_r but not within E_a
- II: An affect minimum within E_a but not within E_r
- III: Affect minima within both E_r and E_a
- IV: Affect minimum "in" E_o (avoid E_r and E_a).

The posited geometric forms of the affect function are illustrated in fig. 3.

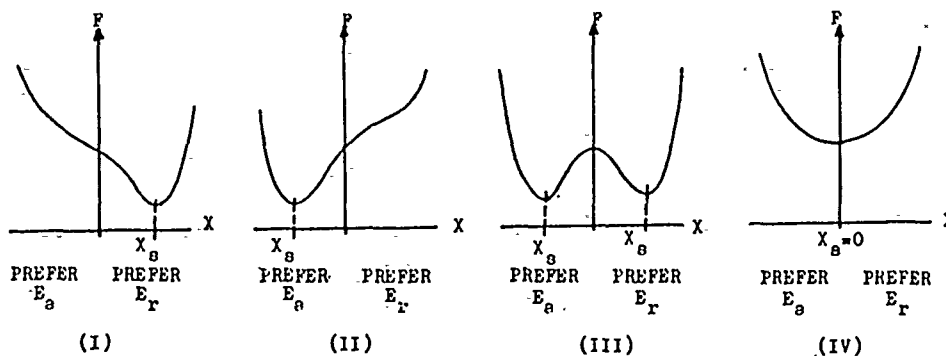


Fig. 3. Schematic representation of qualitative forms of the hypothesized Affect function (F) induced by cognition of "situation" mapped onto potential behavior continuum.

The next developmental step is to prescribe a mathematical model for the affect function (F). At the least, the mathematical model must be able to reproduce the four basic forms shown in fig. 3. Before specifying the math model, let us define the set of variables to be used as determinants of F. There are only five, all of which are implicitly functions of time:

- a - the differential instrumentality of E_r , E_a ($I_r - I_a$)
- b - the magnitude of the choice problem ($I_r + I_a$)
- c - amount of time invested in E_r (tenure)^r
- d - deadline-pressure function (a function of the time remaining on the current contract)
- x - "current" behavioral intention (with respect to staying in or leaving environment E_r).

A mathematical model which will reproduce the qualitative forms shown in fig. 3 (and much more) is:

(1) $F = F_0 + b_1ax + b_2bx^2 + b_3cx^3 + b_4dx^4 + b_6x^6$, where a, b, c, d, and x are as just defined, and F_0 represents the (negative) affect induced by the cognition of having no job at all. It may be noted that equation (1) is equivalent (except for the F_0 term) to the four-parameter, gradient-potential function of the so-called "Butterfly Catastrophe" of Thom (1975). Sheridan (1980) has proposed and tested a two-parameter "cusp" catastrophe model of turnover. And, despite some apparent misinterpretations of the general paradigm, e.g., treatment of the turnover event as a behavior state,

his results tended to support the usefulness of the "catastrophe" approach for turnover modeling. However, the present paper examines an aspect of the problem generally ignored by researchers, including Sheridan.

The current focus is exclusively on the development and interpretation of what has been called the "affect function", F . This is for two reasons: First, if we cannot logically defend the existence of such a function, then we have no reasonable basis for hypothesizing stable states of the presumed dynamic system. Indeed, if such were the case, then we may not have a dynamic system at all (cf. Chillingworth, 1980). Second, suggested applications in the social sciences have largely failed to justify the use of catastrophe theory by any argument, and are accordingly difficult to take seriously.

Regarding the appropriateness of the defined model to describe turnover, an immediate concern is whether or not there is any evidence to support the notion that stated behavioral intentions associate (concurrently in time) with affect minima. Clearly, if there is such an association, then one ought to be able to predict the general manner in which self-reported affect (e.g., worrying about the job decision) would vary with self-reports of behavioral intention. Schneider (1976) surveyed a small group of Navy personnel and asked several questions which are relevant to the current effort. He obtained measures of instrumentality of both military (Navy) and civilian environments with respect to desired life outcomes. His behavioral question asked for the subject's reenlistment intention on a 6-point scale ranging from "definitely intend to reenlist" to "definitely intend to not reenlist. And, of particular interest to the present investigation, Schneider asked his subjects to indicate the frequency with which they thought about whether or not they would make the Navy their career. His 5-point response scale ranged from never think about it to constantly think about it. Demographics included the total length of service (tenure), as well as the time remaining on the current enlistment contract.

Using equation (1) with a set of arbitrary values for the constants F_0 , and the b_i , a set of "affect minima" were calculated for selected points along the cognitive-behavioral continuum. This calculated set of values is plotted in fig. 4 for values of the b -parameter of 2.0, 1.0, and -1.0, c -parameter (tenure) 1.0 (deadline pressure, the d -parameter, was set at zero). For comparison of general form a graph of the mean values of Schneider's respondents to the "think about Navy career" question is superposed (note the displacement).

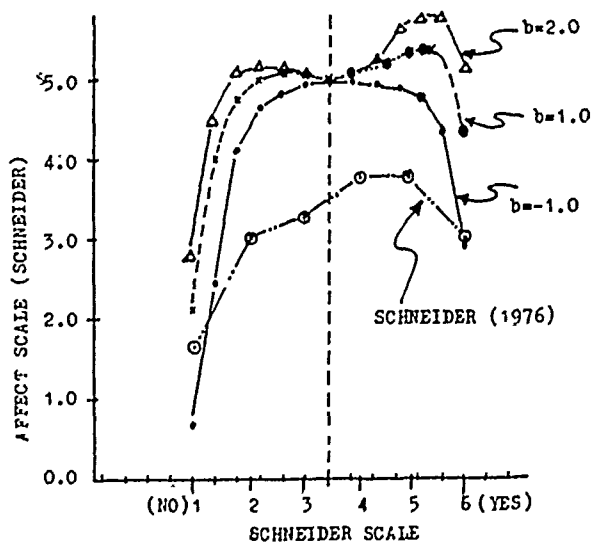


Fig. 4. Comparison of the geometric form of the computed F -function for c -parameter = 1.0, d -parameter = zero, and b -parameter allowed to assume the values 2.0, 1.0, and -1.0. (The computed curves have been displaced upward a small distance (1.5 units) to allow visual comparison of general forms.)

It is readily apparent the curves are qualitatively similar. On the other hand equation (1) expresses F as an explicit function of a, b, c, d , and x . Hence, from a strictly procedural point of view, it is possible to regress the affect surrogate onto estimates of a, b, c, d , and x from the Schneider data set. The result of the subsequent regression analysis accounted for 50% of the variance of F (as defined) ($F_{5, 54} = 10.931$, $P < .0001$). It may be noted that the regression equation thus obtained reflects only the pattern of the posited minima of the F -function, at best. That is, it will approximate the graphs shown in fig. 4, but not those of fig. 3 (the whole of the F -function). Even so, the result is interpreted as suggestive evidence of the existence of the posited affect-minimizing process. Additional theoretical as well as empirical work will be required before stronger conclusions can be made.

References

- Chillingsworth, D.R.J. Structural stability of mathematical models: the role of the catastrophe method. In J.G. Andrews & R.R. McLone (Eds.) Mathematical Modelling. London: Putterworths, 1980.
- Dollard, J., & Miller, N.E. Personality and psychotherapy: An analysis in terms of learning, thinking, and culture. New York: McGraw-Hill, 1950.
- Miller, N.E. Experimental studies of conflict. In J. McV. Hunt (Ed.), Personality and the behavior disorders, Vol I. New York: Ronald, 1944, 431-465.
- Mobley, W.H. Some unanswered questions in turnover and withdrawal research. Academy of Management Review. 1982, 1, 111-116.
- Schneider, J. The "greener grass" phenomena: Differential effects of a work context alternative on organizational participation and withdrawal intentions. Organizational Behavior and Human Performance, 1976, 16, 308-333.
- Sheridan, J.E. A CUSP catastrophe model of employee turnover. Academy of Management Proceedings, 1980, 161-165.
- Tanaka, T. Gels. Scientific American, 1981, 1, 124-138.
- Thom, R. Structural stability and morphogenesis. Reading, Mass.: Benjamin Press, 1975.
- Zeeman, E.C. Catastrophe theory: Selected papers. Reading, Mass.: Addison-Wesley, 1977.

The Greener Grass Phenomenon: Perceptual Distortion
in Individual Decision-Making

James D. Warren
University of Texas at Austin
Austin, TX

Victor H. Appel
University of Texas at Austin
Austin, TX

The Greener Grass Phenomenon: Perceptual Distortion in Individual Decision-Making

Presentations by Watson and Appel have provided background information regarding predecisional reevaluation of alternatives in general and the Schneider model and related study (1976) in particular. Upon the foundation established by Watson and Appel, this presentation will add the following: 1) definition of the Greener Grass Phenomenon; 2) presentation of an extension of the theoretical model suggested by Schneider (1976); 3) placement of the extended model within a broader theoretical context; 4) definition of the Browner Grass Phenomenon; and 5) presentation of the results of further analysis of the Schneider data.

Definition of the Greener Grass Phenomenon

Although Schneider (1976) was the first to use the term Greener Grass Phenomenon in the context of individual decision-making, he did not specifically define the term. For purposes of discussion in this presentation, the Greener Grass Phenomenon (GGP), may be defined as the concomitant decrease in the perceived instrumentality of one's present work environment and increase in the perceived instrumentality of an alternative work environment.

This definition describes what occurs, but does not explain why it occurs. An extension of the model originally postulated by Schneider (1976) provides a means of describing the theoretical mechanics of the GGP.

Extended GGP Model

The constituents of the Extended GGP Model are depicted in Figure 1.

EXTENDED GGP MODEL

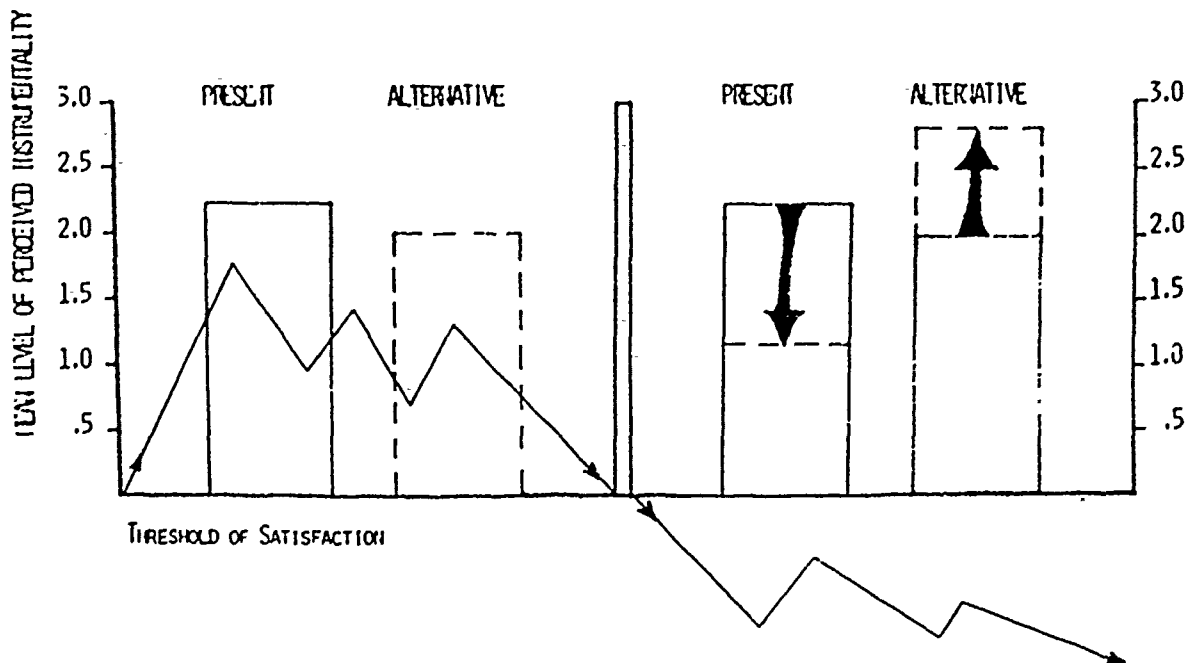


Figure 1. Schematic diagram of the Extended GGP Model (Warren & Appel, 1982).

Inspection of the extended model reveals a collection of variables similar to that of the original Schneider model, but with different functional relationships. The most significant departure from the original model involves the variable of Satisfaction. Whereas the original model depicted Satisfaction as an intermediate and largely superfluous dependent variable, the Extended GGP model places the variable of Satisfaction in the pivotal role of trigger mechanism much in the same manner suggested by Mobley (1982). The extended model further depicts a hypothetical Threshold of Satisfaction above and below which different configurations of instrumentality are observed. As shown in Figure 1, the variables of Satisfaction and the perceived instrumentality of the present (Navy) work environment are hypothesized to covary, with the former acting as the trigger for change in the latter. Consequently, as long as the level of Satisfaction does not fall below the hypothetical Threshold of Satisfaction, the individual's perception of the instrumentality of his/her present work environment will remain relatively high. Even though alternative work environments may be as objectively attractive as the present one, (a) the individual's comparisons of the relative merits of the alternative work environment will be only speculative in nature; and (b) the individual will not seriously entertain the possibility of another work environment as a viable alternative to the present one.

Definition of the Browner Grass Phenomenon

What has been described in the preceding paragraph is the conceptual inverse of the GGP and might aptly be described as the Browner Grass Phenomenon (BGP). The BGP, which may be defined as

the relative parity of instrumentality perceptions of the present and alternative work environments implies a kind of high Satisfaction-contingent cognitive buffer effect. Even though alternative work environments may be as "green" or even "greener" than the present one, the alternatives appear "brown" as long as the level of Satisfaction with the present work environment remains high.

Speculative comparison changes to real comparison, however, once the Threshold of Satisfaction is breached. As the individual's level of Satisfaction drops below the threshold, the level of perceived instrumentality of the present work environment falls and the individual begins to make active comparisons between his/her present work environment and alternatives not seriously considered before. It is at this point that the GGP occurs: an alternative work environment is perceived as having higher instrumentality as a function of (a) decreased Satisfaction, accompanied by (b) a decrease in the perceived instrumentality of the present work environment. Once the Threshold of Satisfaction is breached, the perceived instrumentality of the alternative work environment is artificially inflated, regardless of its objective merit, much as the perceived instrumentality of the present work environment is artificially deflated.

Both sides of the Extended GGP Model portray cognitive/perceptual mechanisms which appear to be (a) Satisfaction-contingent; and which involve (b) a form of perceptual distortion. It is suggested that the Extended GGP Model may have particular utility because it embodies a hypothetical mechanism which describes how and why people leave a given work environment (i.e., the GGP) as well as a hypothetical mechanism which describes how and why people stay in a given work environment (i.e., the BGP).

Theoretical Context

The notion of predecisional reevaluation of alternatives, the very essence of the Extended GGP Model, falls within the purview of a broader theoretical context involving both pre- and post-decisional cognitive activity. Explication of the extended model within that broader context serves to further define by contrast the phenomena already described and also provides a means of making clear conceptual distinctions between pre- and post-decisional cognitive activity. The theoretical context that appears most appropriate for making the necessary distinctions between pre- and post-decisional cognitive activity is that provided by Festinger (1957).

Cognitive dissonance is a term familiar to most students of psychology. As a psychological phenomenon it refers to the post-decisional reevaluation of alternatives so that the chosen alternative is perceived more favorably and the unchosen alternative is perceived more unfavorably. The cognitive/perceptual shift is typically accounted for as the individual's way of rationalizing his/her decision to be "right" or "correct."

What is perhaps not so familiar to students of psychology is the fact that the concept of cognitive dissonance was originally developed as but one component of a larger model. As originally formulated by Festinger, the model included a) predecisional conflict, leading to/resolved by b) a decision, which, in turn, resulted in c) post-decisional dissonance reduction. Prior to the wholesale adoption of cognitive dissonance as the psychological process of interest along that original continuum of cognitive activities, Festinger posed an important theoretical question: are the psychological processes prior to a decision the same as those after a decision has been made?

The question was settled to Festinger's satisfaction as a result of several studies performed by his graduate students (Festinger, 1964). Space limitations preclude detailed examination of the evidence upon which Festinger based his conclusion that the theoretical question had been conclusively answered. Festinger's summary statement, however, is unequivocal.

One may regard the theoretical question as settled. There is a clear and undeniable difference between the cognitive processes that occur during the process of making a decision and those that occur after a decision has been made. Reevaluation of alternatives in the direction of favoring the chosen or disfavoring the rejected alternative is a post-decision phenomenon. While we have as yet cast no light on what cognitive processes do go on during the period of making a decision, at least we know that this one does not occur (1964, p. 31).

By dismissing the possibility of predecisional reevaluation of alternatives, Festinger effectively dismissed the front half of his original model as being theoretically infertile. The present authors contend that a) the evidence upon which the dismissal was made is equivocal at best and in no way supports the broad generalizations drawn therefrom; and b) based on further analysis of the Schneider data, there is sufficient evidence to warrant resurrection of that part of the original Festinger model heretofore condemned to obscurity.

Empirical Evidence

The data furnished by Schneider were subjected to further analysis to determine the level of empirical support for both phenomena (GGP and BGP) ascribed to the Extended GGP Model.

The most likely indicator of either phenomenon was judged to be the criterion measure of behavioral intention to reenlist. This item, depicted in Figure 2, required respondents to furnish a response to the

following question: "Which of the following best describes your reenlistment plans when your current enlistment is completed," along a 6-point scale ranging from "I definitely do not intend to reenlist" to "I definitely intend to reenlist."

INDICATION OF BEHAVIORAL INTENTION

SCHNEIDER(1976), ITEM #16:

WHICH OF THE FOLLOWING BEST DESCRIBES YOUR REENLISTMENT PLANS WHEN YOUR CURRENT ENLISTMENT IS COMPLETED?

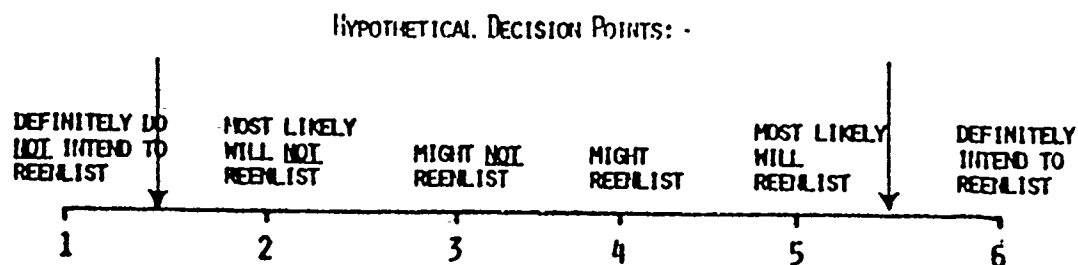


Figure 2. Criterion measure of behavioral intention (Schneider, 1976).

For purposes of analysis, it was assumed that individuals who responded to the item in either extreme had, in fact, already made their decisions. Consequently, the imposition of hypothetical decision points along the scale between scale points 1 and 2 and between scale points 5 and 6 allow a clear distinction between pre- and post-decision phenomena. Since cognitive dissonance is defined as an exclusively post-decisional phenomenon (Festinger, 1957, 1964), the search for evidence of its presence must necessarily be restricted to the area immediately surrounding the hypothetical decision points. Conversely, since the GGP and BGP are hypothesized to be pre-decisional phenomena, the search for evidence of their presence must necessarily be restricted to the middle of the scale.

A preliminary analysis allowed further localization of the theoretical target areas along the scale and, consequently, more specific predictions. The prediction was made that Satisfaction with both the Navy in general and with one's Navy job in particular would be lower for those persons having a Negative Predisposition to Reenlist (i.e., who responded at scale points 1-3 on the behavioral intention measure) than for those having a Positive Predisposition to Reenlist (i.e., who responded at scale points 4-6 on the behavioral intention measure). The results of a oneway ANOVA indicated that there were significant differences in Satisfaction between those subjects with opposite predispositions to reenlist ($p < .001$) and that the differences were in the predicted direction.

Given the evidence for significant differences in Satisfaction between groups of subjects with opposite predispositions to reenlist, the assumption was made that the high Satisfaction evident in those subjects who responded at scale points 4-6 on the behavioral intention measure corresponds to the high Satisfaction condition described in the Extended GGP Model. Since the high Satisfaction condition is the context in which the BGP is hypothesized to occur, scale points 4-5 were specifically identified as the place to look for evidence of the BGP.

Conversely, the assumption was made that the low Satisfaction evident in those subjects who responded at scale points 1-3 on the behavioral intention measure corresponds to the low Satisfaction condition described in the Extended GGP Model. Since the low Satisfaction condition is the context in which the GGP is hypothesized to occur, scale points 2-3 were specifically identified as the place to look for evidence of the GGP.

Scale points 1 and 6 were purposely excluded from BGP and GGP predictions because a clear dissonance reduction effects were predicted at either end of the scale. With regard to perceived instrumentality, the prediction was made that the differences between the mean levels of Navy and civilian instrumentality between scale points 1 and 2 would be both significant and in the direction consistent with the reenlistment predisposition, i.e., between scale points 1 and 2 the mean level of perceived Navy instrumentality was predicted to decrease and the mean level of perceived civilian instrumentality was predicted to increase. Similar predictions, consistent with the reenlistment predisposition, were made for the other end of the scale, i.e., between scale points 5 and 6, the mean level of perceived Navy instrumentality was predicted to increase and the mean level of perceived civilian instrumentality was predicted to decrease.

A oneway ANOVA with orthogonal contrasts was performed to test the predictions specified. As depicted in Figure 3, the results were mixed and somewhat surprising.

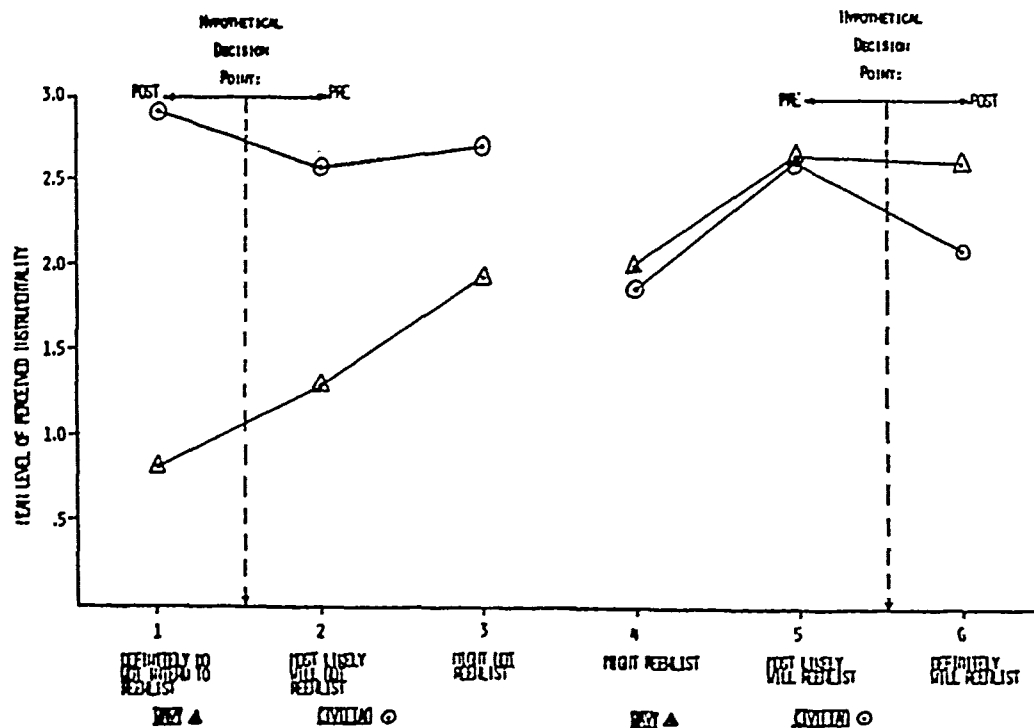


Figure 3. Mean instrumentality levels across the behavioral intention scale.

For example, the predicted dissonance reduction effect occurred at one end of the scale but not at the other. The mean differences in perceived Navy and civilian instrumentality between scale points 1 and 2 were significant at the .01 level of significance. The mean differences in the perceived instrumentalities between scale points 5 and 6, however, failed to achieve significance at conventional levels.

The hypothesized GGP effect received equivocal support. Whereas there was no significant difference in civilian instrumentality between scale points 2 and 3, the difference in Navy instrumentality was also nonsignificant ($p < .24$) but moving in the predicted direction.

The hypothesized BGP effect received strong support. Between scale points 4 and 5, both Navy and civilian instrumentality moved in the same direction and the differences between the two instrumentalities were not statistically significant.

Discussion and Implications

The results reported, while somewhat equivocal, provide evidence that may be interpreted as encouraging, given the exploratory nature of the research and the level of development of the theoretical model. It should also be noted that the analyses reported do not constitute a full test of the Extended GGP Model because Satisfaction was not systematically varied. Such an analysis has been performed and will be reported elsewhere (Warren & Appel, 1982, in press).

Of all of the phenomena for which predictions were made, the Browner Grass Phenomenon (BGP), the relative parity of perceived Navy and civilian instrumentality under high Satisfaction conditions, received the strongest support. Evidence for the Greener Grass Phenomenon (GGP), the concomitant increase in perceived civilian instrumentality and decrease in perceived Navy instrumentality under low Satisfaction conditions, was equivocal but in the predicted direction.

The mixed results with regard to the predicted dissonance reduction effects were perhaps the most puzzling. Evidence of a strong dissonance reduction effect at one end of the scale (i.e., Negative Predisposition to Reenlist) and its absence at the other (i.e., Positive Predisposition to Reenlist) introduces the intriguing possibility that there may be different decision processes at either end of the scale. More specifically, these results suggest the possibility that it may be easier to decide "No" than "Yes" with regard to reenlistment. If this is the case, the implications for retention policies and practices could be significant. There may exist elements of the Navy work environment that are perceived to be so negative that they provide the basis for a relatively easy "No" decision that is manifested by the clear dissonance reduction effect between scale points 1 and 2. Conversely, elements of the Navy work environment that might support a "Yes" reenlistment decision may not be nearly as salient. The encouraging implication of this hypothetical scenario is that the Navy may be in a position to influence both the positive and negative factors, if they exist, to attain a more desirable retention rate.

References

- Festinger, L. A theory of cognitive dissonance. Stanford, Ca.: Stanford University Press, 1957.
- _____. Conflict, decision, and dissonance. Stanford, Ca.: Stanford University Press, 1964.
- Mobley, W.H. Some unanswered questions in turnover and withdrawal research. Academy of Management Review, 1982, 7, 111-116.
- Schneider, J. The "Greener Grass" phenomenon: Differential effects of a work context alternative on organizational participation and withdrawal intentions. Organizational Behavior and Human Performance, 1976, 16, 308-333.

AIR FORCE WIVES AND OFFICERS WIVES CLUBS

Marianne Pierce, M.A.
University of Maryland

Vince Luchsinger, Brig. Gen. USAFR
University of Baltimore

The role of the service wife has been of consistent interest to the Armed Forces community for a considerable length of time. The contribution of the wife to the career of the service member is often related to Officer Wives Club activity. Given the changing role of women in society, the nature and perceptions of Air Force wives in Officers Wives Clubs are subjects of scrutiny for social scientists (McCubbin, Dahl and Hunter, 1976; Carr, Orthner, and Brown, 1980).

This study investigated the opinions of over 200 officers wives at a relatively small Air Force Base. A stratified random sample, with follow-up mailings and telephone calls, provided 117 usable responses dealing with the experiences and activities of the Air Force wives, including Wives Club functions.

Activities The wives were queried about their involvement in different types of activities in the community. The responses provided the following:

Attribute	Responses	% Yes
Volunteer Work	117	21
Either Student or Employed	115	68
Childcare Duties	116	48
Membership in Other Military Organizations	117	30
Membership in Sorority or Education Clubs	116	9
Membership in Religious Organizations	116	17
Membership in Professional Organizations	116	18
Consider Self Active in any Membership	116	48
(Active in Officer Wives Club Only: 14%)		
(Active in OWC and in Other Clubs: 16%)		
(Active in Other Clubs, Inactive in OWC: 70%)		

The responding wives were then asked to make comments about their participation and satisfaction in regard to Wives Club activity. Given that not every respondent had actual contact with an Officers Wives Club, comments and perceptions were solicited in various areas of activity categorized by positive or negative nature of the comment.

A total of 160 positive comments were made about various categories of OWC activity. The categories, with numbers of responses and some sample statements are as follows:

Charitable (33): "The money raised for the needy is wonderful"

Social (90): "It offers an opportunity to meet other AF wives".
"New friendships can be begun at each new base"

Educational (35): "We learned a lot about entertainment productions"
"We planned luncheons, fashion shows, and a dance show"
"It's a good way to learn about the military"

Economic (2): "It can help your husband to be known by the brass"
"It is a tool to help me gain visibility for my profession"

A total of 219 negative comments were derived from respondents. While social comments were the most numerous positive comments, negative comments were led by inflexibility, social, and efficiency. The categories with numbers and sample statements are:

Charitable (10): "I think OWC doesn't do enough for its own members"

Social (52): "It's not a very supportive group; too rank conscious"
"Cliques and little groups form and alienate others"
"There was not too much mingling"

Educational (8): "I found the program to be boring; poorly done"
"OWC could offer more intellectually stimulating programs"

Economic (28): "OWC is for volunteers; I prefer to be compensated"
"It costs to belong, as well as for baby sitters; costs add up"

Inflexibility (51): "The Board is a closed group to outsiders"
"The Club hasn't changed as much as the times have"
"There aren't enough evening sessions for working women"
"I want to be myself; I feel I'm being watched to see if I fit in"

Inefficiency (43): "Meetings are too long"
"Communication is often poor or lacking altogether"
"I've not been contacted by OWC or I would have joined"
"After the Orientation process, I heard little about the organization"

Other (27): "Too busy"
"OWC takes too much time; once you agree to help, it's always expected of you"

The responding group provided a wide range of responses, with the negative opinions somewhat outnumbering the positive responses. The nature and range of the responses is not unlike other volunteer membership groups.

Many expected relationships did not turn out to be statistically significant. For example, there was expected to be a relationship between husband's rank and whether or not a wife was employed outside the home or attended school. Intuitively, one might think younger wives would be most likely to be engaged in activities outside the home such as work or school. No such relationship was found. Wives of Field Grade Officers seemed as likely as younger wives to be attending school or working outside the home. No relationship was found between husband's rank and whether or not a woman does volunteer work. That is, the older wives were as unlikely to perform volunteer work as the younger wives in this particular sample. On the other hand, some significant relationships did turn up using chi square at the .05 level.

More military wives are working than in the past, as most people would guess. Sixty-eight percent of survey respondents was found between OWC membership and whether or not a wife works and/or goes to school. Of those women who neither work or go to school, a higher proportion of them are not OWC members. Among those women who work, go to school, or both, the higher proportion of respondents are OWC members. This runs contrary to what might be expected, even though authors such as McCubbin & Marsden (1979) remind us that today's military family has become transformed from the traditional framework. As women become involved in outside activities such as school or work, one might expect to find fewer of them as OWC members. We found the reverse to be statistically significant, as the following table illustrates.

	Non-Members	OWC Members	Percentage Totals
Neither Work Nor Attend School	59.5	40.5	100
Work Or Attend School Or Both	30.8	69.2	100

Grade Comparisons A statistically significant relationship was found between OWC membership and rank of respondent's husband. As rank increased, a higher proportion of respondents tended not to belong to OWC. This relationship is shown by the following percentages.

	Lieutenants	Captains	Field Grade
Non-Member	31.58	55.22	57.1
OWC Member	68.42	44.78	42.86
TOTALS	100.00	100.00	100.00

Several explanations are possible. It may be, as many wives stated, that younger wives join to learn more about the military and the military life-style. Another possible explanation is hinted at in comments of respondents. That is, some younger wives fear that not belonging will possibly harm their husband's long-range chances for promotion, and this maybe less of a concern for wives of Field Grade Officers.

A significant relationship appeared between husband's rank and whether or not one lives on base. Lieutenants and Captains tend to live on base more often than field Grade Officers. This relationship is likely a reflection of the lower pay levels of Junior Officers as compared to Field Grade Officers.

OWC vs. Non - OWC Member Comparison A series of t-tests (statistical tests of significance) were run to compare the responses of OWC members and non-members on each positive and negative category. With a few exceptions, the responses of members and non-members were not significantly different. For example, members and non-members did not differ in the number and types of clubs/organizations to which they belonged. Members and non-members did not differ in the extent to which they consider themselves to be active in OWC or other club memberships. Members and non-members were also similar in the degree of positiveness they assigned to their positive comments. A similarity was also found between members and non-members in their statement and ratings of negative comments/incidents. For example, in the category of "inflexibility" 30 members and 19 non-members had contributed comments, which made up the category. Non-members' average rating of negativeness was 3.7, while OWC members' average rating of negativeness was 4.2; however, this did not turn out to be a statistically significant or reliable difference. The more important information to note may be the overall rating of both members and non-members who listed "inflexibility" type comments. The combined average rating of the negativeness for this particular category of comments is near 4.0 on a 5-point scale, indicating that both members and non-members who perceive OWC as inflexible also believe it to be a highly negative attribute.

A significant difference was found in how members and non-members rated the importance of their negative economic responses. OWC members attributed more importance to their negative economic comments than did non-members, although, both members and non-members rated negative economic attributes of OWC as very important. On the five-point scale, the average ratings were over 4.0 for both OWC members and non-members. The overall ratings of this attribute when the members and non-members are combined reached 4.5, which is extremely high.

A statistically significant difference was found between OWC members and non-members on ratings of negativeness of perceived OWC inefficiency. Comments/incidents in this category were considered less important by members than by non-members. This may mean that members are more tolerant of organizational inefficiencies or that inefficiencies are more a problem of OWC image than a problem in reality.

In summary, the make-up, satisfactions, and activities of Air Force wives are more complex than ever. More studies are underway to perform in-depth analyses of a wide range of officer wives related to quality of life programs.

Another series of t-tests were performed to examine possible differences in responses of those who work or go to school versus those who do neither. Again, few significant differences were found. In most cases, the responses of women who work or go to school were like the comments of women who neither work nor go to school.

There was a difference in the ratings of the two groups of respondents on the negative economic variable. Wives who work or go to school rated the economic variable more negatively than did wives who neither work nor go to school. Once again, however, we see both groups reacting very strongly to perceived negative economic aspects of OWC. Averages are 4.0 or higher on a scale of 1 to 5. In trying to interpret these figures it should be kept in mind what the category contains. This category contains comments related to the cost of membership plus babysitters, the cost of luncheons, the amount of unpaid time required to be an active member, the fact that the husband's career may be advanced or hindered by the wife's membership/participation. All we can be sure of is that members perceive the economic variables more negatively than do non-members.

	Member	Non-Member	Combined Overall Average
Average Rating of Negativeness of Economic Variables	4.8	4.0	4.4

Finally, it should be noted that all 31 respondents reporting educational variables as positive aspects of OWC were off-base people. This unanimous response pattern is notable. Inflexibility and inefficiency type comments were commonly made by women who are working and/or attending school. This suggests that educational functions should be stressed to off-base people and those new to the military. Wives who are working or students might be solicited for help in streamlining the efficiency of OWC. Comments about economic liabilities of OWC suggests that it may be feasible to explore possibilities of reducing dues to those who can and will devote more time to charitable/organizational work and making up the deficit from those who can better afford to contribute money instead of time.

REFERENCES CITED

Carr, Richard; Orthner, D.K.; and Brown, R.J., Living and Family Patterns in the Air Force, Air University Review, 1980.

McCubbin, H.I.; Dahl, B.B.; and Hunter, E.J., Families in the Military System Beverly Hills, CA: Sage Publications, 1976

McCubbin, H.I. and Marsden, M.A. The Military Family and the Changing Military Profession. in The Changing World of the American Military, Frank Margiotta (ed.) Boulder, CO: Westview Press, 1979

AIR FORCE FAMILY SUPPORT CENTERS

Bobby Donohue
Los Angeles AFS Family Support Center
V. P. Luchsinger, Brig.Gen. USAFR
University of Baltimore

Quality of Life studies in the Air Force have indicated that the condition of the family of the service member is a very strong factor in morale and personal productivity. Nowhere is this more important than in urban areas with extensive commuting, expensive housing where available, and the stress of a complex environment. The Los Angeles area, home of the USAF Space Division, is one such location. In a hotbed of technological innovation and production, heavy financial, energy, and personal demands are made on personnel assigned to Los Angeles Air Force Station.

Enter Family Support Centers, which were proposed by the 1980 Air Force Conference on Families to serve as a focal point in addressing family needs in Air Force communities. Fielding of Family Support Centers (FSC's) is an important new family initiative supported by Air Force leaders, program managers, and Air Force families.

FSC's will attempt to:

1. Focus on "Air Force Family" needs; single, married and their families.
2. Provide a one-stop single point of service or referral.
3. Enhance existing programs by focusing attention on and raising the awareness of the needs of Air Force families.
4. Facilitate two-way communication between the family and the Air Force.

Need. FSC's are needed for a variety of reasons. Presently, support services for Air Force families are fragmented and vary base to base. Most support programs are staffed and have resources to support active duty military family members on a "space available" basis, or not at all. The number of volunteers available to work in existing programs has declined in recent years. Still, 67% of all Air Force members have families.

Families have come to be considered a vital mission support system, since a positive correlation has been established between family attitudes and the retention/productivity of Air Force members. With this impetus, the Air Force leadership looked toward a new umbrella program to work the family support problem.

Status. The Air Force Family Matters (AFFAM) was established to design and implement a program for improving the quality and quantity of family support.

Four prototype centers were opened in late FY 81 (Kadena AB, Okinawa; Bitburg AB, Germany; Moody AFB, Georgia, and Travis AFB, California). In addition, three MAJCOM supported Family Support Centers were programmed for Los Angeles AFS, Zweibrucken, and Upper Heyford.

In 1982, eight new FSC starts are anticipated, one per MAJCOM. In 1983, twelve new FSC starts are scheduled; one per MAJCOM, one at the Air Force Academy, plus two others. If evaluations show that Family Support Centers are effective, a total of 124 FSC installations are forecast by 1987. This represents an ambitious and significant investment by the Air Force in the well-being of the Air Force "family system".

Staffing. A prominent feature of the Family Support Center is the use of trained and experienced staff. Experience has shown this to be a key aspect in the success of family support systems.

Directors of the base centers will be experienced and professionally qualified at the GS 11-13 level. With the benefit of Air Force institutional training, the Director will operate as manager, educator, and clinician.

NCOIC's will be assigned as chief of management and administration. The NCOIC of the Los Angeles AFS Center is an E-8 with extensive experience as a first sergeant. Two to three program technicians will be assigned at the GS 7-10 level. They will possess professional and technical qualifications. Additional administrative and support personnel will be available depending upon the size of the base.

Administration. The Family Support Center is intended to be mission-responsive. It will be directly responsible to the Base Commander as a tool for the Commander to respond to the needs of his/her people. The FSC will not duplicate existing services but link families with existing resources.

Base FSC programs will be developed according to certain guidelines:

1. Needs assessments will ascertain family requirements, programs, and resources. Contractor assistance will be provided at the prototype bases.
2. Programs will be tailored to, and prioritized, according to local needs.
3. Advisory boards will be made up of families and base service providers to act as advocates and communication links with base and area families of qualified personnel.

Experience thus far indicates that the programs are on track. For example, at Los Angeles AFS, staffing and facilities are secured. The existence of command support and resources has been instrumental in the progress to date. This has been critical in an area such as Los Angeles with many social, economic, and geographic factors to complicate family living.

Family Support Centers will provide programs and services in eight categories that cover the span of areas which impact on family life.

Planned Family Support Center Programs are:

1 o Referral/Coordination

- family media information
- commander/first sergeant spouse leadership course
- program to train supervisors to recognize family problems
- spouse networks
- volunteer programs
- babysitting cooperatives
- base/community family services referral program
- family newsletter
- family column in base newspaper
- referral for community health providers (doctors/dentists)

2 o Relocation

- relocation stress training
- INTRO/OUTRO family sponsorship
- real estate fundamentals
- base information kit library
- foreign culture orientation for families enroute overseas
- orientation counseling

3 o Family Separation

- family support groups
- ombudsman program
- community support during separation

4 o Spouse/family employment

- job information
- skills bank
- vocational education/information

- academic education/information
- reentry education
- "how to find a job" training
- resume preparation
- career counseling
- education assistance, grants, loans, scholarships

5 o Families in crisis

- financial crisis
- domestic violence
- death of a family member
- family therapy
- acting out children
- coping with child/adolescent development
- help children cope with separation/divorce
- sexual abuse
- rape
- coping with stress
- crisis intervention
- juvenile offenders program

6 o Financial management

- household finances
- home economics
- consumer education
- financial aid programs
- health benefits assistance

7 o Family development education

- parenting
- marriage enrichment

- couples communication
- family communication
- personal growth
- child development
- family development
- assertiveness training
- resources for exceptional children
- family recognition
- family recreation
- pre-marriage training
- marriage counseling
- creative parenting
- value clarification
- positive attitude clinics
- family wellness program

8 o Special needs families

- retirement transition
- single parent support groups
- military couple cooperatives
- American Culture for foreign born spouses
- citizenship training
- foster parents/grandparents programs
- retired activities programs

In summary, the Family Support Center will be a tool for commanders to identify and help meet changing needs of Air Force members and their families. FSC represents a full program development based on evaluation of prototype programs now under way.

JOB SATISFACTION AMONG U.S. AIR FORCE ACADEMY HOMEMAKERS

David B. Porter
Sharon J. Porter

ABSTRACT

Hackman and Oldham's Job Diagnostic Survey was used to analyze the job of homemaker. Results showed that the job of homemaker was slightly richer than the national norm. The major area of dissatisfaction was the lack of opportunities for homemakers to interact with one another. The homemaker's age and years of marriage showed a strong negative correlation with the homemaker's view of how "enriched" their job was, while homemaking help from the husband had a significantly positive impact on the perceived job enrichment.

Dissatisfied homemakers spent slightly more time outside the home, particularly doing volunteer work, than those who were more satisfied. The results of the survey suggest that homemaker dissatisfaction was not strongly correlated with the number of hours homemakers spent in paid employment or attending school.

INTRODUCTION

Although based on Herzberg's original theory of job enrichment, J. Richard Hackman and Greg R. Oldham's "job characteristics approach" expands the boundaries of application to include failures as well as successes (Miner: 1980, p259). Their approach suggests that certain core job characteristics impact on workers' psychological states which in turn directly influence their attitudes and performance on the job. Five core job characteristics (skill variety, task identity, task significance, autonomy and feedback) combine to determine three critical psychological states which influence certain outcomes such as worker motivation and job satisfaction (Hackman and Oldham: 1980, p77). When all five characteristics are high the model predicts that a worker's internal work motivation will also be high. When one or more of these characteristics is low the model predicts that the worker's motivation will be less. Very high scores on some characteristics can compensate for lower scores on other characteristics. Thus, although each job has its own unique profile (set of scores on the five core job characteristics), very different types of jobs may be equally motivating.

The instrument developed by Hackman and Oldham to measure these variables is the Job Diagnostic Survey (JDS). This survey has been used to study 876 different jobs and establish national norms for job characteristics (Hackman and Oldham: 1980, p105). It identifies which characteristics of a job are strong and which need to be improved. By improving job

characteristics identified as lacking in motivation, the job is enriched and the quality of worker performance should increase while rates of absenteeism and worker turnover should decrease (Korman: 1977, p140). This effect, however, is moderated by the individual worker's growth need strength and their satisfaction with working conditions, pay and benefits, and supervision. Hackman and Oldham's job enrichment theory is based on the premise that workers have different levels of growth needs and are not all equally satisfied with supervision, job security, and the social atmosphere at work (Hackman & Oldham: 1980, 115-120). To the extent that workers have high growth need strength and are satisfied with these other aspects of their jobs, Hackman and Oldham's model suggests that job enrichment will increase both worker motivation and job satisfaction (Mitchell: 1978, p398).

The job of homemaker has been a part of society since society began, but it has not received nearly as much attention as most other jobs, in spite of the fact that it is the primary type of job for a very large number of people. The job of homemaker has been praised by some as though the people who undertook this job were martyrs or saints, leading one to the conclusion that being a homemaker was more a "calling" than an occupation. It has also been criticized and called a "non-job". It has been suggested that individuals who undertake the job of homemaker were "copping-out", and not making a meaningful contribution to society. This paper is taking neither position, but merely studying the job of homemaking as any other job and comparing it to other jobs by using Hackman and Oldham's Job Diagnostic Survey.

Specifically our approach was to identify which of the five core job characteristics increased homemaker satisfaction and which core characteristics caused dissatisfaction in comparison to the national norms reported by Hackman and Oldham (1980, p317). We also attempted to find out what effect age, length of marriage, number and ages of children, and the amount of help received from the husband and children had on the homemaker's evaluation of their five core job characteristics.

Method

Hackman and Oldham's Job Diagnostic Survey was distributed in the Douglas Valley housing area of the Air Force Academy. Fifty questionnaires were distributed to randomly selected clusters of homes. The purpose of the survey was explained to each homemaker who was home (distribution was made Sunday afternoon and evening to maximize the number of people home). However, for approximately ten homes where the homemaker was not home, the survey was left on the front door. Thirty-four (68 percent) of the surveys were completed and returned in the self addressed envelopes provided. It should be noted that this sample might differ from other homemakers in the civilian community or even homemakers on other military bases. Most military members are "specially selected" to either teach or conduct military training for Air Force Cadets.

Demographic characteristics of the homemakers in the sample are shown in Table 1. Although we did not ask respondents to identify their sex on the survey, none of the homemakers we talked to were male nor did any of the comment sheets returned with the survey suggest that a male had

responded as the homemaker. Therefore we assumed that our sample was entirely female. Table 1 shows that the average homemaker responding to the survey was 30-34 years old, had been married 10-12 years, had two children between the ages of six and twelve and that both the husband and children did minimal chores on a regular basis. Table 2 shows that the average homemaker spent a total of approximately 35 hours outside the home in activities such as volunteer work, paid employment, school, and other activities.

The data provided by the survey was analyzed using the Statistical Package for Social Sciences (SPSS) BREAKDOWN program. The resulting means and standard deviations were then compared with national norms and the differences were noted. Correlations (r) yielded by SPSS BREAKDOWN's One Way Analysis of Variance (ANOVA) and significance of linear relationships were used to determine those demographic characteristics which had a significant effect on the homemaker's core job characteristics. ANOVA was also used to determine the effect of affective outcomes, context satisfactions and individual growth need strength on hours spent by the homemaker outside the home.

RESULTS and DISCUSSION

The means and standard deviations for each of the variables are compared with the national norms in Table 3. The final column of this table shows the net differences in the job of homemaker and other jobs (positive values indicate areas in which homemaker means were higher than the national norms). The results of the survey showed that the job of homemaker was slightly lower in skill variety, higher in task identity and the same in task significance. Homemakers reported having slightly more autonomy and receiving more feedback than workers in other jobs, but also reported having much less opportunity to deal with others. The first five core job characteristics combine to determine the job's motivational potential score (MPS). (See Miner: 1977, p235.) The homemaker's MPS of 144 compares favorably to the national norm of 128. It should be noted that the national norm of 128 includes all types of jobs including professional and highly technical occupations. The MPS advantage for homemakers is even greater when compared to the average MPS for the jobs women traditionally enter [i.e. clerical (MPS = 106) and processing (MPS = 105)] (Hackman and Oldham: 1980, p317).

The affective outcomes from the survey showed the job of homemaker as being nearly the same as the national averages for all jobs. This indicates that the homemakers surveyed were generally satisfied with their jobs as homemakers.

The context satisfactions or extrinsic factors show that homemakers are generally satisfied with their job security (+.4), pay and benefits (0), and supervision (-.1) but that they score well below the national norm (-.8) in social satisfaction. This observation is consistent with the earlier finding that the job of homemaker did not provide much opportunity for dealing with others (see Table 3). It is also consistent with Ingraham and Manning's finding of "pervasive expressions of isolation

DEMOGRAPHIC CHARACTERISTICS OF
RESPONDENT TO HOMEMAKER SURVEY

VARIABLES	LEVELS/NUMBER (PERCENTAGE)					
	25-29	30-34	35-40	40-44		
Age	6(18%)	17(50%)	9(26%)	2(6%)		
Years of Marriage	1-3	4-6	7-9	10-12	13-15	16-18 Over 19
	2(6%)	5(15%)	5(15%)	12(35%)	7(21%)	2(6%) 1(3%)
Number of Children	0	1	2	3	Over 4	
	3(9%)	6(18%)	17(50%)	4(12%)	4(12%)	
Age of Oldest Child	0-1	2-5	6-12	13-18		
	0	7(23%)	18(58%)	6(19%)		
Age of Youngest Child	9(29%)	7(23%)	13(42%)	2(6%)		
Amount of Homemaking Help from Husband	Does Almost Everything	Share Chores Equally	Does Minimal Chores If Asked	Does A Few Things	None	
	1(3%)	7(21%)	14(41%)	10(29%)	2(6%)	
Children	2(6%)	11(32%)	13(38%)	5(15%)	3(9%)	

Table 1

HOMEMAKER HOURS SPENT OUTSIDE
THE HOME

Activity	Estimated Average Number Of Hours Spent Per Week				
	0	1-5	6-15	16-35	Over 35
Volunteer Work	16(47%)	14(41%)	1(3%)	3(9%)	0
Paid Employment	14(41%)	2(6%)	7(21%)	7(21%)	4(12%)
School	28(82%)	2(6%)	2(6%)	2(6%)	0
Other Activities	1(3%)	9(26%)	20(59%)	4(12%)	0
TOTAL	0	2(6%)	8(24%)	12(35%)	12(35%)

Table 2

COMPARISON OF HOMEMAKER JDS DATA
WITH NATIONAL NORMS

Variables	Homemaker		National Norms ^a		Net Differences
	X	SD	X	SD	
Job Characteristics					
Skill Variety	4.0	1.7	4.7	1.2	-.7
Task Identity	5.6	1.5	4.7	1.3	.9
Task Significance	5.5	1.4	5.5	1.2	0
Autonomy	5.2	1.2	4.9	1.3	.3
Feedback from Job	5.0	1.4	4.9	1.2	.1
Feedback from Others	4.7	1.7	4.1	1.4	.6
Dealing with Others	4.1	1.8	5.6	1.0	-1.5
Motivating Potential Score(MPS)	144	78	128	73	16
Affective Outcomes					
General Satisfaction	4.5	1.2	4.7	1.2	-.2
Growth Satisfaction	4.9	1.4	4.8	1.2	.1
Internal Work Motive	5.5	1.2	5.6	.8	-.1
Context Satisfaction					
Job Security	5.3	1.7	4.9	1.3	.4
Pay	4.3	2.1	4.3	1.4	0
Social	4.5	1.4	5.3	.9	-.8
Supervision	4.8	1.4	4.9	1.5	-.1
Growth Need Strength	5.7	1.2	5.0	.7	.7

^aNote: The data in National Norms columns is from Hackman and Oldham(1980)

Table 3

and lack of belonging" among military homemakers (1981:5). Homemakers have less interaction with others than the average worker and seem to miss it.

Their score for individual growth need strength is .7 above the national norm. This might help explain why homemakers derive only average satisfaction from a job that is above average in motivational potential (Hackman & Oldham: 1980, p118).

The correlations between demographic variables and job characteristics are shown in Table 4. The homemaker's age and number of years married had a negative correlation with most job characteristics. These correlations were significant at the .05 level for both autonomy and overall MPS. We suspect that this is due to changing social values rather than "job deterioration". It is important to note here that this was not a longitudinal study comparing job characteristics of a group of homemakers over a period of time but rather a cross sectional study of homemakers of different ages.

The age and number of children had no significant effect on the homemaker's perception of the "richness" of their jobs. Although the amount of help from children was generally positively correlated with job scores, none of the relationships were statistically significant.

The most significant findings among the demographic characteristics were found in the amount of help from the husband. The more the husband helped the greater the amount of autonomy the homemaker felt and the more they felt that their jobs as homemakers provided an opportunity for them to deal with others. Help from the husband was also significantly correlated with homemaker MPS.

Correlations between affective outcomes, context satisfactions and growth need strength are shown in Table 5. As general satisfaction, growth satisfaction, and internal work motivation decreased, the amount of hours the homemaker spent doing volunteer work increased significantly. This is a strong indication that volunteer work is the homemaker's equivalent of absenteeism, especially when these negative correlations are combined with the significantly negative correlations between volunteer hours and satisfaction with supervision. Among the 34 respondents to the survey there were three individuals who spent over 20 hours doing volunteer work and all three were very dissatisfied with their jobs as homemakers. Although the above correlations are significant they are based primarily on the responses of these three individuals.

The results of the survey also suggest that homemakers do not seek paid employment due to their dissatisfaction with their jobs as homemakers. Paid employment was not significantly correlated with growth need strength, context satisfactions, or any of the affective outcomes.

The amount of job security was significantly negative correlated with the number of hours the homemaker spent at school. The less job security the homemaker felt the more hours were spent on education. Time spent in school was not significantly correlated with any other variables.

CORRELATIONS (R) BETWEEN
DEMOGRAPHIC CHARACTERISTICS
AND JOB CHARACTERISTICS

DEMOGRAPHIC CHARACTERISTICS	SKILL VARIETY	TASK IDENTITY	TASK SIGNIFICANCE	AUTONOMY	FEEDBACK FROM THE JOB	FEEDBACK FROM OTHERS	DEALING WITH OTHERS	MPS
Age	-.09	-.17	-.30	-.41*	-.24	-.17	-.19	-.37*
Years of Marriage	-.18	-.15	-.17	-.43*	-.21	-.16	-.26	-.36*
Number of Children	-.05	.04	-.06	-.06	-.05	-.13	.12	-.02
Age of Oldest Child	-.11	.05	-.19	-.21	.05	.09	0	-.12
Age of Youngest Child	.05	.13	-.16	-.15	.08	-.02	.06	-.03
Homemaking Help from Husband	.13	.21	-.02	.44**	.26	.22	.59**	.35*
Homemaking Help from Children	.12	.06	.12	-.18	.19	.15	.11	.08

*p < .05

**p < .01

Table 4

CORRELATIONS OF AFFECTIVE OUTCOMES,
CONTEXT SATISFACTIONS AND GROWTH NEED STRENGTH
WITH HOURS SPENT OUTSIDE THE HOME

	HOURS SPENT OUTSIDE THE HOME				
	Volunteer Work	Paid Employ	School	Other Activity	TOTAL
AFFECTIVE OUTCOMES					
General Satisfaction	-.39*	-.01	-.21	-.04	-.25
Growth Satisfaction	-.46**	0	.01	-.16	-.23
Internal Work Motivation	-.40*	-.11	-.13	.26	-.19
CONTEXT SATISFACTIONS					
Job Security	-.10	.23	-.37*	-.13	-.04
Pay	-.26	-.08	.05	-.19	-.21
Supervision	-.36*	.14	.06	-.17	-.07
Social	-.29	.26	.18	-.27	.05
GROWTH-NEED STRENGTH	-.21	0	.01	-.03	-.10

* p < .05

** p < .01

Table 5

The only other areas which approached significance were the correlations between time spent in other activities with internal work motivation and social satisfaction. Homemakers who were highly motivated as well as those seeking interaction with others tended to spend more time outside the home.

CONCLUSION

The job of homemaker is relatively rich when compared to other jobs, particularly when compared to the national norms for processing and clerical jobs. Task identity (the extent to which a person completes whole job from beginning to end and is able to see the outcome of his and her work) was the core job characteristic which reflected the homemaker's greatest relative advantage.

One drawback to the job of homemaker, however, was noted in the area of context satisfactions. This was the homemaker's satisfaction with her opportunity for social interaction. As the age of the homemaker and the years of marriage increased the amount of autonomy homemakers felt decreased. This seems to be associated with changing social values and views of marriage: the traditional view that the female homemaker was solely responsible for all housekeeping and child-rearing tasks compared to the emerging view that these tasks should be equitably shared between both marriage partners. Homemakers' dissatisfaction with social opportunities might also help to explain the lure of alternative communal life styles which became popular in the late 1960's. The extent to which the husband helped the homemaker had a direct and strongly positive influence on the amount of autonomy and social interaction the homemaker reported. The more the husband helped the more satisfying the job. This might be because the spouses were able to interact while completing some tasks together.

One of the most interesting results of this survey was the relationship between satisfaction and the hours spent outside the home. The amount of volunteer work reflected a strong negative correlation with all affective outcomes and seems to be the homemaker's equivalent of industrial absenteeism. Another interesting correlation was the amount of time spent going to school and the homemaker's dissatisfaction with her job security. Some homemakers seem to be attempting to create some personal security through increased education to compensate for the lack of security they feel in their jobs as homemakers.

REFERENCES

- Hackman, R.J. and Oldham, G.R. Work Redesign; Reading, Massachusetts: Addison-Wesley Publishing Company, Inc., 1980.
- Ingraham, L.H. and Manning, F.J. "Cohesion: Who Needs It, What Is It and How Do We Get It to Them?", Military Review, 1981, No. 6, 3-12.
- Korman, A.K., Organizational Behavior; Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1977.
- Miner, J.B. Theories of Organizational Behavior. Hinsdale, Illinois: Dryden Press, 1980.
- Mitchell, T.R. People In Organizations; Understanding Their Behavior. New York: McGraw-Hill Inc., 1978.

DUAL-CAREER FAMILY NEGOTIATIONS
AMONG U.S.A.F. ACADEMY CADETS

Frank R. Wood
and
Leona Flores
U.S. Air Force Academy

Abstract

In-depth interviews with 15 Air Force Academy women cadets and their intended spouses, who were either cadets or officers, reveal clear relationships between sex, career choice, and primary marriage or career identity.

Traditional marriages are oriented toward promoting occupational achievement on the part of the husband aided to great measure by the facilitative emotional and domestic support of the wife. In contemporary marriages, this precept is challenged by an emerging pattern of dual-career marriages in which the wife is not willing to suppress her own wishes for occupational achievement. Dual-career marriages are a growing concern of the Air Force and the situation of cadets who are marrying other cadets or officers provides a unique opportunity to study this marriage form at the earliest stages.

This research examines the premarriage negotiations of Air Force Academy cadets who are about to enter both their careers and marriages simultaneously. These negotiations reveal normative orientations and expectations internal to these dual-career marriages and describe clear relationships between marriage or professional identities, sex, and career choice.

METHOD

The findings in this study are based on in-depth interviews with 15 Air Force Academy women cadets and their intended spouses who were either cadets themselves or already in the Air Force. The problems of dual-career marriage were especially salient to these couples because they were interviewed just prior to graduation and were currently working with the personnel system to obtain joint assignments in the Air Force. The interviews were unstructured, conducted with each individual spouse, and involved questions merely designed to get the respondent to talk about the difficulties of beginning both a career and marriage together. Selection of the couples was guided by

the evolving theory that career choice reflected primary identity. Eventually the sample consisted of: five couples who were both pursuing flying careers; five couples in which one spouse would be flying and the other would enter a support specialty; and five couples in which both were entering support fields.

FINDINGS

In the process of describing the negotiations required to establish a career and marriage simultaneously, the respondents described the relationships summarized in Table 1. In this model, career choice is categorized as "flying" or "not flying" and identity reflects a major emphasis on "marriage" or "career." In general, sex and career choice are related to the traditionality of the of the partnership and inversely related to the requirement for negotiation. In other words, the more non-traditional arrangements required more negotiation. A full appreciation of how these variables interact to determine primary identities can be gained through an analysis of specific negotiations associated with each couple type.

TABLE 1
DETERMINANTS OF PREMARITAL NEGOTIATION POWER

Males			Females	
Career Choice	Primary Identity	N	Primary Identity	Career Choice
Flying	Career	5	Career	Flying
Non-flying	Marriage	0	Career	Flying
Flying	Career	5	Marriage	Non-flying
Non-flying	Marriage	5	Marriage	Non-flying

The DUAL FLYING COUPLES are a unique group whose situation requires the most negotiation. For both men and women in this group, professional identities are well established. Flying was viewed as a "once in a lifetime opportunity" and a main reason they attended the Air Force Academy rather than a civilian college. Four of the five men had clearly established plans to become fighter pilots. For the woman, this non-traditional career choice represented a strong commitment to a professional identity.

The problem for dual flying couples is that flying careers tend to involve inflexible schedules, limited assignment possibilities and a high potential for separation. Although all pilots are initially trained at the same schools, specialized follow-on training locks a pilot into a specific weapon system and constrains future assignments to particular bases where the aircraft of specialization are located. Joint assignment to any particular base, then, requires both partners to be trained in the same or at least compatible aircraft. This may seem a small matter; but it is complicated by current restrictions of women in combat weapon systems. Few non-combat aircraft assignments are compatible with the highly prestigious and desired fighter aircraft assignments. If they do get married, the options for this couple-type are limited: either the male will have to accept less than a fighter assignment, the female will have to give up a flying career for a more flexible support career, or both must accept separation in the early years of marriage.

The importance of the professional identity in this couple type is highlighted by the comments these couples offer about their pre-marriage negotiations. In all cases, professional identity was more important than marriage identity to both members. These comments were typical of this group:

He wasn't used to dating someone who wants a career... He assumed that I wouldn't want to fly and I told him that flying was the only reason I was here at the Academy... Then I told him it was either me or fighters. (female)

We already discussed this separation before we decided to get engaged. I definitely want a fighter and we will experience a three-year separation if it is the only way. (male)

The non-traditionality of this arrangement appears in the plans which these couples were making about having children. Most couples planned to wait until the woman had finished her flying career. Only one couple, four years older than the others (a medical constraint), planned children sooner, but not until the woman's follow-on training was completed.

The FLYING/NON-FLYING COUPLES are interesting because they provide further explanation of the relationships between key issues. Logically, this group could have two forms: a flying male and a non-flying female; or a flying female and a non-flying male. The latter is a very non-traditional arrangement as it gives the female higher occupational prestige than the male. Not surprisingly, no cases of this type were found, probably because when

couples are given the choice, they opt for the more traditional relationship in which the male flies and the female does not. In part, this hypothesis was substantiated by a couple in which both partners could fly, but opted for the more traditional arrangement in which the women would not fly. The husband had this to say about the less traditional option they declined:

I don't want my wife to fly... If she flew, she'd have a big ego and we only need one person to have a big ego in this relationship.

Traditional orientation, then, is seen as offering a more stable relationship and given a free choice, traditionality seems to be the point of equilibrium.

Unlike the dual flying couples, who faced separation later in their relationship, the problem for the fly/non-fly couples is separation immediately after marriage. The flying spouse is required to attend flying training which lasts one year and the non-flying spouse must attend a technical training school which lasts seven months to a year and one half. After this initial separation, the couple probably will be assigned together, assuming the non-flying spouse has chosen a compatible career field. The optimum solution for this couple type is to choose a career field for the non-flying spouse which requires the least separation initially and is very flexible thereafter.

This process of "building in flexibility" has some interesting consequences for the primary identities of each sex in this couple type -- especially the non-flying spouse (usually the woman). Like the dual flying couples the flying spouse in this group has a stronger career identity. Many talked about delaying the marriage until after finishing pilot training because the flying spouse viewed doing well in training as initially more important than the marriage. Also, the flying spouse could make few concessions because the flying career is so constrained and inflexible. The result is a situation in which all the flexibility must be created by concessions of the non-flying partner. This is problematic because deferring to the inflexibility of the flying spouse, the non-flying partner is put into the role of having to be more concerned with family identity than professional identity. Further, differential concern with family matters in these early stages may set a pattern which casts the non-flying spouse as "less interested in professional development" which is likely to affect later negotiations as to whose professional career is more important.

These relationships were highlighted by this couple-type's comments about their negotiations. Typically, the non-flying spouse, the woman, described having few choices which would provide the needed flexibility. One of the women illustrated her dilemma as follows:

I am going to be an air traffic controller. I chose this because it was compatible with flying. I didn't have much choice.... either aircraft maintenance, communications or air traffic controller. I didn't care much for the others, so I chose the controller job.

The one case in this group that deviated from this pattern of the non-flying spouse placing secondary importance on professional identity was a woman who had an opportunity to attend medical school. Like the flyers, she saw this as a "once in a lifetime opportunity" which was more important than marriage. She described the situation this way:

Medicine is more important to me than to be with him; I don't want to give up this opportunity. I would regret it later, I think.

Beyond this unique opportunity, the woman in this case confirms the normal relationships found among members of this couple type. In a later comment about contingency plans she stated:

If I wash out of medical school, I will go for a job that is most compatible with the plane he will be flying.

The traditionality of this couple type was emphasized in their negotiations about children. The majority of the males in this group wanted their wives to resign from the Air Force and raise children. Others implied that if the wives had to have careers, "minimal careers" in support fields were acceptable. Thus, the primacy of family identity for the non-flying spouse is reinforced by the flying spouse.

The DUAL NON-FLYING COUPLES seemed to have the optimum situation for the marriage. Careers in this case are very flexible and easy to manipulate for joint assignment. Conflicts between the career and the marriage are usually resolved in favor of the marriage. Typical comments regarding conflicts reflected a primary concern for the marriage:

If staying together means changing my career, then I am willing to sacrifice my career... I probably can find something else to do that will be fine. If we constantly get split up, we'll just leave the Air Force.

It is clear that both partners in this couple-type place the greatest importance on marriage rather than on career identities. In fact, this is the only, couple-type in which the males are generally supportive of the female careers.

Their support is illustrated by these comments:

I realized flying wasn't the way to go if we wanted to be together so I gave up flying to be with her.

She wants to pursue a law career... That's alright with me... I'll even take care of the kids.

In general, these comments describe a couple type characterized by a shared married identity, relatively balanced power and few critical negotiations.

Plans about children supported this contention. While couples of this type generally planned to wait a few years to have children, both partners were confident the wife could "juggle the responsibilities of career and motherhood." One male even predicted he would be the one who raised the children.

CONCLUSION

This model summarizes the relationships of key factors affecting the pre-marital negotiations of cadet couples who are about to begin both their careers and their married life. While it in no way represents relationships to be found among dual career couples in the Air Force, it does constitute the starting point for these relationships. Perhaps at this early stage in the development of a marriage, the influences of sex and career choice on primary identity are especially salient because both professional and marriage identities are in the formative stages of development.

In this sample, a clear pattern was revealed. Sex and the flexibility associated with the career choice determined the relative importance of the marriage or career identity. Career choice affects identity in two ways: the inflexibility of one partner requires the other to create flexibility by placing less importance on professional concerns; and the non-traditionality of cross-sex career choices may generate pressure for identity choices to align with standard sex roles. Identity also seems to be a function of the relative importance of the career choice. In cases where the career choice is seen as a "once in a lifetime opportunity," the career identity tends to overshadow the marriage identity. Otherwise, the marriage identity dominates. Early identities are important to later negotiations about dual-career conflicts because differential concern with the marriage generates vulnerability. While these correlates of primary identities should be of concern to all, they should be of special concern to the Air Force. Where factors operate to diminish the importance placed on professional identities, the members more concerned with the marriage identity (usually the woman) tend to be less committed to a career.

THE CONTRIBUTION OF INTELLECTUAL ABILITY AND ORGANIZATIONAL EXPERIENCE TO LEADERSHIP PERFORMANCE

FRED E. FIEDLER

UNIVERSITY OF WASHINGTON

Contemporary leadership theory basically ignores the effects which intellectual abilities and job-relevant experience might have on leadership performance. Such formulations as Theory X and Y (McGregor, 1967), the Managerial Grid (Blake and Mouton, 1965); Path Goal Theory (House and Mitchell 1974), Vroom and Yetton's managerial decision model (1973), or the Contingency Model (1967), do not consider intellectual abilities or job relevant variables.

On the other hand, empirical evidence linking intelligence and leadership performance has been very meager (Stogdill, 1974). The median correlations between intelligence tests and performance measures have been in the range of .22 to .28, and such experience measures as time in the organization and time in leadership jobs have been unrelated to leadership performance. One review of 13 studies reported a median correlation between time in the organization, or in supervisory or managerial positions and leadership performance of -.12 (Fiedler, 1970) and subsequent studies have yielded similar results. There is, therefore, little reason to believe that these cognitive functions contribute greatly to organizational performance.

These empirical findings also seem borne out in everyday life. The most intelligent leaders are not necessarily most effective; and university professors are certainly not known for their managerial prowess.

A very similar conclusion emerges in studies relating the leader's experience to performance. Again we find that the most senior managers are not always most effective, and that young and relatively inexperienced leaders often outperform their more experienced colleagues. Several well-known historical examples of this type are the 17 year old Joan of Arc, the 19 year old Alexander the Great, and William Pitt, who at the age of 24 became one of England's most distinguished prime ministers.

It is very difficult to believe, however, that intelligence and experience are irrelevant in leadership or managerial positions. This notion is certainly not accepted by the practitioner. Personnel managers and executive "head hunters" usually do not go after dull and inexperienced managers; and assessment centers invariably consider a candidate's intellectual abilities and work history.

The curious inconsistency of the research findings and the basis on which leaders and executives are actually recruited and selected, obviously requires an explanation. It is hard to understand why intelligence and general ability tests are such poor predictors of leadership performance. It is equally difficult to believe that managers learn nothing of value from having spent time in an organization of leadership position.

The leader must surely get to know the organization's subculture, its role

system, standards, expectations, mores and taboos, as well as the idiosyncracies of co-workers. He or she must also learn something about the technical aspects of the job: how to troubleshoot a machine, to whom to complain about a delayed payroll check, how to get around an irksome rule of the company. And we assume that leaders can learn to motivate people, to discipline, and to cope with an eyeball-to-eyeball confrontation.

And yet, leadership researchers have paid very little attention to the role of experience in leadership jobs. In fact, Stogdill's "Handbook of Leadership" (1981) does not even list the term in the index, nor do most management texts, e.g., Campbell, Dunnette, Lawler, and Weick's (1979), or the more recent text by Rambo (1982). In effect, we know next to nothing about how experience is acquired, what is learned, and how it is applied.

One possible explanation for the apparently conflicting evidence from empirical studies and from institutional practice suggests itself. Intellectual processes and job relevant knowledge (experience) might contribute to leadership performance only under certain conditions. In fact, under some circumstances the more intelligent and experienced leaders might well perform less well. This is exemplified by the "know-it-all" who threatens his boss, the "smart aleck" who does not bother to listen to others, and the ultra-conservative oldtimer who learned nothing and forgot nothing since he joined the organization.

One important factor in the effective utilization of intellectual abilities is likely to be the interpersonal stress a person might have with the immediate superior. Everyone knows of a teacher or a boss who is able to reduce subordinates to a state of near imbecility where they can do nothing right, and where all the appropriate answers come one half hour after the interview. Stress of this type may well inhibit the ability of a leader to think logically or to solve difficult problems. On the other hand, the experienced old-timer has gone through the same drill over and over again, and has developed set ways of reacting. By falling back on previously learned ways of dealing with recurrent problems, the experienced person is able to cope more effectively with certain job requirements in stressful situations.

Experienced leaders also are likely to know how the organization really works --the short-cuts and ways of untangling red tape and how to stay out of trouble. We can expect, therefore, that the experienced leader will perform better than the less experienced leader when relations with the boss are strained.

We usually infer the contribution of a person's ability to a task by the magnitude of the correlation between the ability score and the individual's task performance. For example, we assume that mathematical ability contributes to performance in engineering school because of the relatively high correlation between mathematical ability test scores and grade point average.

The contribution of intellectual or general ability to leadership performance may likewise be inferred from the magnitude of the correlation between leaders' general ability test scores and leadership performance. The identical argument can, of course, be made for inferring the contribution of experience to leadership performance.

By noting the conditions under which the correlations between general ability and performance, or organizational experience and performance, are high, we can

identify the factors which determine how intellectual abilities and experience contribute to leadership performance. This paper discusses the contribution of two important factors, namely, stress with the boss, and critical requirements of the task.

The Effect of Stressful Relations with the Superior.

With the exception of immediate family, the boss is usually the most important single individual in one's life. Without the boss' support or endorsement, advancement to higher rank or to better jobs is almost impossible. And even when we leave the organization, the recommendation of our immediate supervisor may, for many years to come, determine the path our career will take. The boss' praise and reward can make life highly satisfying and rewarding, while the boss' disapprobation, not to mention petty aggravations, can make life on and off the job a source of acute pain and misery. This is especially so in military organizations where even one lukewarm efficiency report can permanently stunt an officer's military career. Most of the data I shall here review come from military organizations.

The first study in this series was based on 158 army infantry squad leaders, that is, men with from about 4 to 8 years of army service, who supervise the work and training of 11 enlisted men. Their job calls for knowledge of the various basic infantrymen's skills, and the maintenance of army discipline and standards in their units. These men report to a platoon leader, a relatively young second lieutenant for whom this is usually the first army assignment, and to the platoon sergeant with 6 to 12 years of army experience.

Squad leaders' general ability (AB) was measured with the Army's General Classification Test, experience was defined as time in service (TIS); stress with boss (SB) was measured by a semantic differential scale bounded by the adjectives "very stressful--not at all stressful." Performance (PRF) was measured on a 16 item scale completed by the platoon leader and platoon sergeant. The scale asked raters to mark the degree to which the person to be rated exceeded, met, or failed to meet various requirements, e.g., "how well he organizes his group", "the way he handles the job when the demands are extra heavy," etc.

The various scores (even SB and PRF) were uncorrelated. It was, therefore, possible to break the sample into those with high and with low SB, and to determine the degree which AB and TIS correlated with performance. As the results of these analyses showed (Table 1), AB and PERF correlated only when SB was low. TIS did not correlate with PERF. A further division of the group into those with very low or very high SB (± 1 SD of the mean) yielded a somewhat higher relationship between AB and PRF under low boss stress, as well as a significant correlation between TIS and PRF when stress with boss was very high.

Table 1

A subsequent study (Fiedler, Potter, Zais & Knowlton, 1979) involved 45 company commanders and their first sergeants, as well as 47 staff officers of 9 army battalions. Most of the officers held captain's rank, although a substantial number of first lieutenants were staff officers. Intelligenece was here measured with the Wonderlic Personnel Test, a 12 minute, 50-item scale which has been highly correlated with other ability tests, (e.g., the intelligence scale of the SRA Multi-Aptitude Test, U.S. Bureau of Personnel Management, 1965); experience

was defined as time in service, stress with boss by means of short scales, and performance was evaluated by the battalion commander and battalion executive officer, using the Bons scale from the squad leader study. Line and staff officers had nearly identical mean INT, TIS, and PRF scores. Staff officers had somewhat lower SB scores which did not differ significantly from those of line officers.

Another study (Potter and Fiedler, 1981) dealt with 130 officers, petty officers and civilian employees in staff positions at the managerial level. Intelligence and experience, as well as performance were measured as in the study of company commanders and staff officers. Stress with boss was assessed by means of a 34 item scale on which staff personnel rated the degree of stress caused by various boss behaviors, e.g., "He acts unfriendly and unapproachable", "He does not tell me what he expects of me".

In addition, respondents also rated the structure of the task, and they indicated the proportion of time they devoted to each of 10 important staff functions (routine paper work, advising on policy issues, etc.), and they rated the stress from the job itself.

We again divided the group on the basis of stress with boss, and then correlated general ability and performance for those with low and with high boss stress. Figure 1 summarizes the data for the army and coast guard studies. This figure shows the magnitude of the correlation (in z-scores) on the vertical axis, and the low and the high boss stress condition on the horizontal axis. Each point on the graph thus indicates the degree to which general ability or organizational experience (time in service) contributed to leadership or staff officer performance. (Correlations were computed separately for coast guard personnel with intellectually demanding and less demanding tasks. These will be discussed below.)

Figure 1

As can be seen, the results generally support the findings of the squad leader study. In each case, the correlations between ability and performance were positive when SB was low, but near zero or negative when SB was high. Correlations between time in service and performance were positive only when SB was high.

It is of particular interest to note, however, that we obtained low positive correlations between staff officers' general ability and performance when SB was low, but relatively high negative correlations (-.40 to -.51) when SB was high. Thus, while intellectual ability neither assisted nor detracted from the staff officer's performance when stress with superior was low, high SB substantially decreased the staff officer's utilization of his intellectual ability and enhanced that of his organizational experience.

Several further studies have been conducted. One study by Borden (1980) involved 442 officers and noncommissioned officers (NCOs) of an army infantry division. The composition of the sample is shown on Table 2. Measures were essentially the same as in previous studies, i.e., Wonderlic, time in service, and Bons scale measured, respectively general ability, organizational experience, and performance, and a scale similar to that used in the coast guard study assessed stress with boss as well as other types of stress to which we shall return later.

Table 2

Figure 2 shows the average of the correlations between performance and general

ability (Wonderlic) and time in service when each of the subsamples was divided at the mean into a lower middle and an upper third on the SB scale. As can be seen the findings supported those of previous studies. It is of particular interest to note, however, that general ability and organizational experience again have opposite effects on performance and that the correlation curves are almost mirror images of each other. In the low stress condition, in which general ability contributes to performance, organizational experience contributes, general ability tends to detract.

One plausible explanation of this finding is the well-known fact that we perform habitual behavior less well when we consciously think about it (e.g., thinking of your hand while playing tennis). On the other hand, someone who has performed the same job for many years is less likely to think of a better way of doing it than someone who comes to it new. These findings also indicate, of course, that intellectual or general abilities contribute to leadership performance only under certain conditions, i.e., here when the relationship with the immediate superior is relatively free of stress. This means that our ability to predict leadership performance from ability tests or indices of experience and organizational knowledge will be considerably enhanced by taking into account of the specific conditions in which the leader will perform the task, or teaching the leader how to develop situations in which his or her particular abilities are used to best advantage.

Critical Task Requirements and the Contribution of Intellectual Abilities and Experience.

It seems obvious upon brief reflections that leaders who have the particular abilities which the task requires should perform better than those who do not happen to have these particular abilities. Thus, for example, an intellectually demanding task requires more intellectual ability than an intellectually undemanding task. It should, therefore, be performed better by the intellectually more able leaders. Intellectually more able leaders would not necessarily be more effective on tasks which demand, for example, interpersonal skills and sensitivity, or athletic prowess.

It will be recalled that we had asked each subject in the Coast Guard study to indicate the time devoted to each of ten commonly performed staff functions. These included, for example, routine paper work, inspecting field units, and advising on policy issues. We had also asked a separate group of coast guard officers to rank each of these ten functions on the intellectual demand each functions made of the job holder.

Staff personnel who rated themselves as relatively high (above the median) on time devoted to a particular job function were divided into a high and a low SB group, and correlations were then computed between their performance and their ability and organizational experience scores.

The results of these analyses are shown on Table 3, with job functions listed in the order of rated intellectual demand. The table also shows median correlations for the five functions high in intellectual demand and those low in intellectual demand.

Table 3

As expected, general or intellectual ability contributed more to performance

under low stress when the task was intellectually demanding than when it was relatively undemanding. However, intellectual abilities were also more detrimental under high SB when the tasks are intellectually more demanding. As expected, correlations between Time in Service and performance were unaffected by the intellectual demands of the task.

The effects of stress and critical job requirements are especially apparent in Borden's study of 9 army infantry battalions. As will be recalled, data were collected on platoon sergeants, platoon leaders, first sergeants, company commanders, executive officers, and battalion staff officers. These different positions obviously entail tasks which differ on several dimensions.

A group of 13 officers and senior NCOs made ratings of these six jobs on several dimensions which they considered to be relevant. We shall here consider two, namely, the degree to which the task is intellectually demanding, and the degree to which it requires tactful management of the leader's relations with his immediate superior: e.g., to keep his boss out of trouble, to make the unit more effective, or to make his own life easier.

As mentioned earlier, Borden had obtained various measures of interpersonal and job stress. One important source of boss stress turned out to be "Overcontrol" by the boss, e.g., meddling and autocratic, dictatorial behavior. The major job stress factor turned out to be "Uncertainty", e.g., poor management, job and role ambiguity.

It is now possible to determine the effects of the task and of boss stress on the contribution of intellectual abilities and experience to leadership performance.

- (a) We first order the five leadership jobs on the degree to which they make intellectual demands and to which they require management of relations with the boss.
- (b) Leaders are then divided into those reporting low or high Overcontrol Stress.
- (c) We can then plot the appropriate correlations between general ability and organizational experience with performance.

Figure 4 shows that the correlations between the Wonderlic Test Score and leadership performance increase with the increasing demands for intellectual effort of the job. The greater the job's critical requirement for intellectual effort, the greater the contribution of the leader's intellectual ability to performance--but only for leaders with low boss stress.

Figure 4

Figure 5 shows the analogous correlations when jobs are ordered on the leader's need to "manage" difficult interpersonal relations with the boss, tactfully guiding his superior for the benefit of the organization or other purposes.

Intellectual abilities apparently are not useful for meeting this particular task requirement. Rather, the leader must know the organization, the subculture, customs and traditions, which guides the interactions among people. This knowledge is acquired in the course of time, as can be seen on Figure 5. Those who have been in the army longer (high TIS) perform leadership jobs of that type better than those who have had less army experience. In contrast to findings related to intellectual ability, job experience is utilized more effectively when stress with the boss is high rather than low.

Figure 5

These findings suggest that a "gut approach", based on an intimate knowledge rather than carefully planned action, assists the leader to get along. This again suggests that experience teaches us the sub-culture of the organization, and that an intellectual, problem solving approach does not contribute to performance in jobs of this type. This may be reflected in the commonly held view that the calculating, coldly manipulative person may not be suited for the management of interpersonal relations.

It is also apparent from the above correlations that the relatively unstructured tasks, that is, those of company commanders or platoon leaders, which require the use of discretionary power, benefit from the leader's intellectual abilities; interpersonally demanding tasks, e.g., of first sergeants, benefit from experience and knowledge of the organization. These general intellectual abilities are detrimental to performance in jobs which require shrewdness and the ability to manage difficult interpersonal relations.

We have also found, however, that the highly experienced manager, who feels comfortable with his boss, is likely to develop overconfidence and contempt for jobs that are highly routinized and structured. In these he is then likely to perform less well than his relatively inexperienced counterpart who takes the work more seriously.

In conclusion, I have here discussed two important factors which determine, at least in part, the role of general or intellectual ability and of organizational experience. It goes without saying that the measures we have used are at best primitive, and that more precise indices of ability and experience should provide better understanding of the contribution of cognitive abilities and knowledge to leadership performance. Moreover, although I have not discussed these factors in the present paper, we have ample evidence that the leader's personality and motivation also need to be considered in a more comprehensive theory which would describe the role of cognitive abilities and knowledge in leadership performance.

BIBLIOGRAPHY

- Bass, B.M. , Stogdill's Handbook of Leadership New York: Free Press, 1981.
- Blake, R.R., and Mouton, J.S. Managerial Grid Houston: Gulf 1964.
- Borden, D.F., Leader superior stress, personality, job satisfaction, and performance. Another look at the interrelationships of same old constructs in the modern large bureaucracy. Doctoral dissertation, University of Washington, 1980.
- Campbell, J.P., Dunnette, M.D., Lawler, E.E. & Weick, K.E. Managerial behavior, performance, and effectiveness. New York: McGraw-Hill, 1970.
- Fiedler, F.E. A theory of leadership effectiveness New York: McGraw-Hill, 1967.
- Fiedler, F.E. Leadership experience and leadership performance-another hypothesis shot to hell. Organizational Behavior and Human Performance, 1970, 5, 1-14.
- Fiedler, F.E., Potter, E.H., Zais, M.M., & Knowlton, W.A. Jr. Organizational stress and the use and misuse of managerial intelligence and experience. Journal of Applied Psychology 1979, 64, 634-647.
- House, R., and Mitchell, T.R. Path-goal theory of leadership. Journal of Contemporary Business, 1974, 3, no. 4, 81-97.
- McGregor, D. The professional manager. New York: McGraw-Hill, 1967.
- Rambo, W.W. Work and Organizational Behavior New York: Holt, Reinhart and Winston, 1981.
- Stodgill, R.M. The handbook of leadership: a survey of theory and research New York: Free Press, 1974.
- U.S. Employment Service, U.S.Department of Labor. Correlation of the General Aptitude Test Battery, Form B-1002B with the Wonderlic Personnel Test, Form B. Technical Report, U.S. Department of Labor, October, 1964.
- Vroom, V.H., & Yetton, P.W. Leadership and decision-making. Pittsburgh: University of Pittsburgh Press, 1973.

TABLE 1
CORRELATIONS BETWEEN INTELLIGENCE AND EXPERIENCE WITH
TASK PERFORMANCE FOR LEADERS AND STAFF MEMBERS
REPORTING VERY HIGH OR VERY LOW STRESS WITH THEIR SUPERIORS

LEADER'S <u>STRESS WITH SUPERIOR</u>	<u>IQ</u>	LEADER <u>EXPERIENCE</u>
VERY LOW	.51**	.09
VERY HIGH	-.01	.40*

STAFF MEMBER'S <u>STRESS WITH SUPERIOR</u> †	<u>IQ</u>	STAFF MEMBER'S <u>EXPERIENCE</u>
LOW	.73**	-.07
HIGH	-.43**	.44*

† STAFF MEMBERS WITH LESS THAN 10 YEARS OF EXPERIENCE

*p < .05

**p < .01

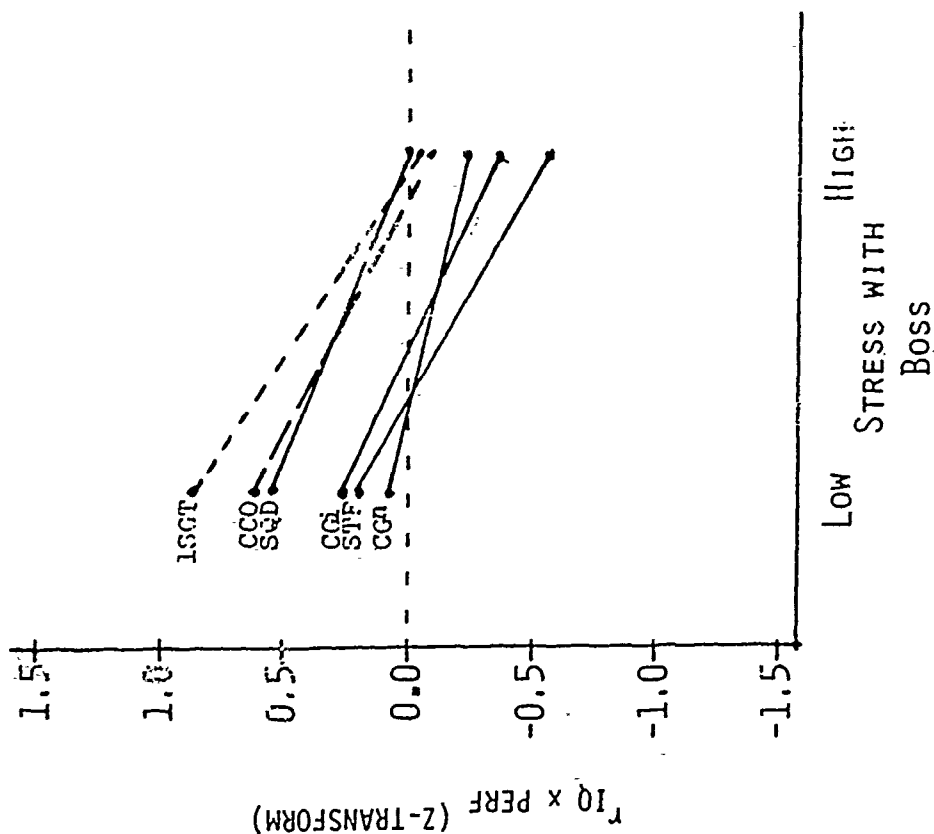
Table 2

BATTALION STAFF OFFICERS	(BSO)	N=43
COMPANY COMMANDERS	(CCO)	N=4
COMPANY EXECUTIVE OFFICERS	(CXO)	N=34
FIRST SERGEANT	(FSG)	N=42
PLATOON LEADER	(PLT)	N=106
PLATOON SERGEANT	(PSG)	N=163

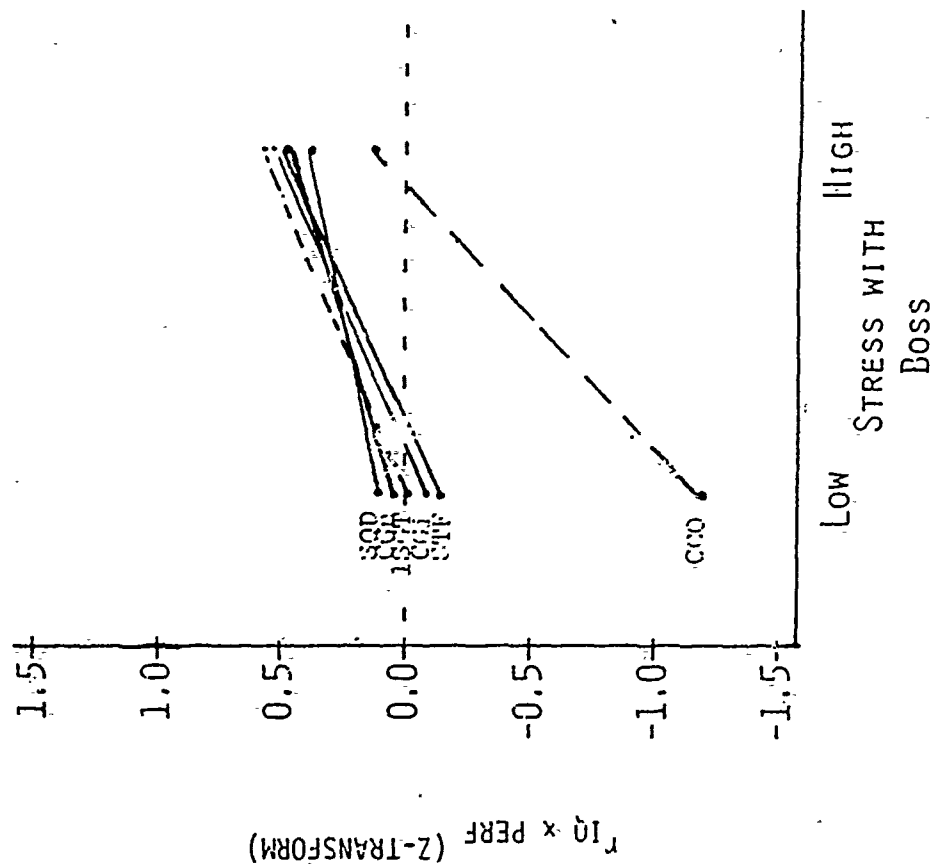
COMPOSITION OF THE SAMPLE OF SUBJECTS FROM THE BORDEN STUDY (1980). THREE LETTER DESIGNATION OF THE JOB IS GIVEN IN PARENTHESES. NUMBER OF SUBJECTS WITHIN EACH JOB CATEGORY IS SHOWN IN THE RIGHT COLUMN

FIGURE 1

INTELLIGENCE X PERFORMANCE

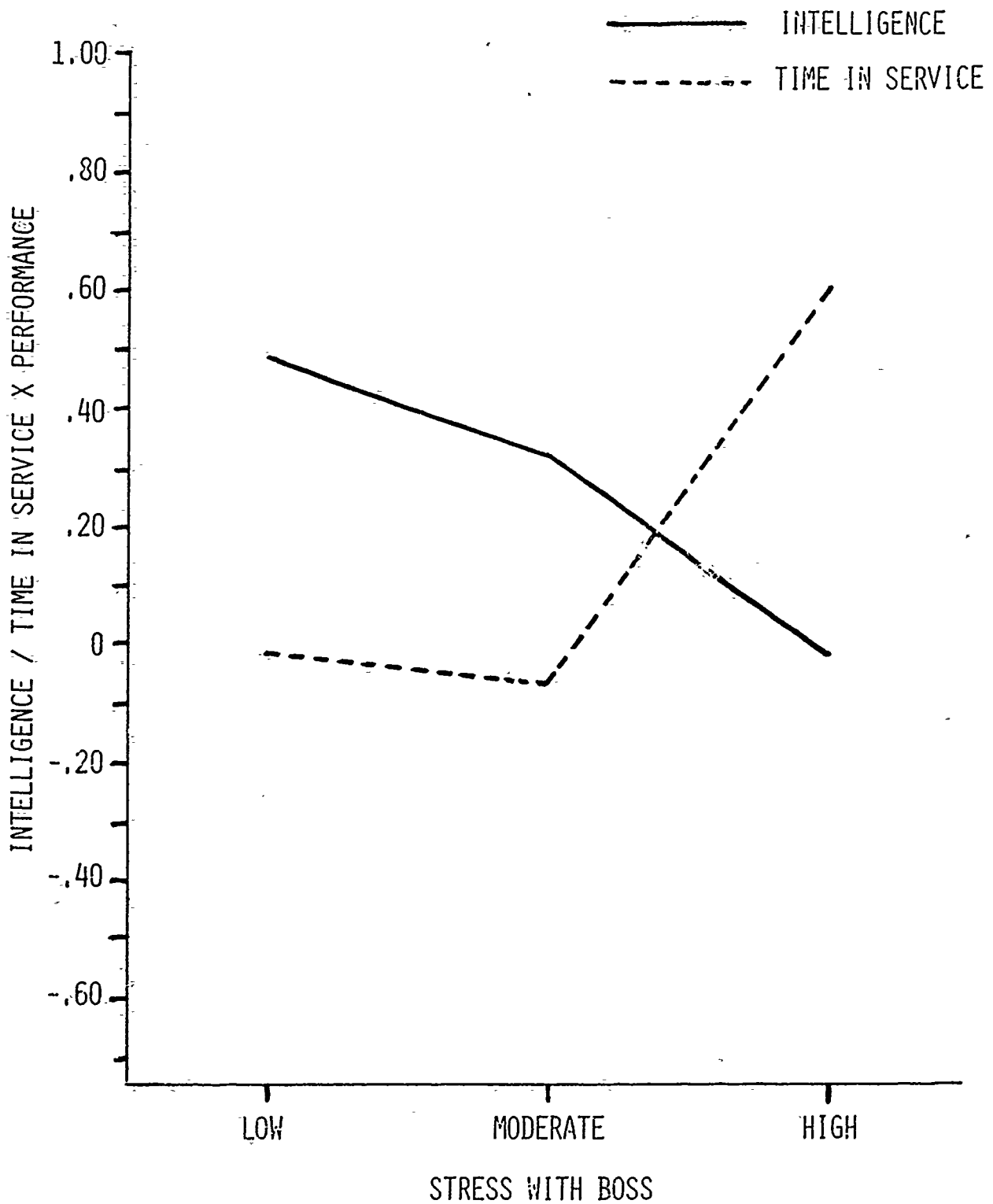


EXPERIENCE X PERFORMANCE



CORRELATIONS OF PERFORMANCE WITH INTELLIGENCE AND EXPERIENCE OF MILITARY PERSONNEL
FROM VARIOUS ORGANIZATIONS UNDER CONDITIONS OF LOW AND HIGH STRESS WITH SUPERIOR

FIGURE 2



MEAN CORRELATIONS BETWEEN INTELLIGENCE TEST SCORES AND RATED LEADERSHIP PERFORMANCE AND TIME IN SERVICE AND RATED LEADERSHIP PERFORMANCE UNDER LOW, MODERATE, AND HIGH STRESS WITH BOSS FOR OFFICERS AND NON-COMMISSIONED OFFICERS OF NINE INFANTRY BATTALIONS.

Table 13
CORRELATIONS OF PERFORMANCE AND INTELLIGENCE
AND PERFORMANCE AND EXPERIENCE
UNDER LOW AND HIGH BOSS STRESS

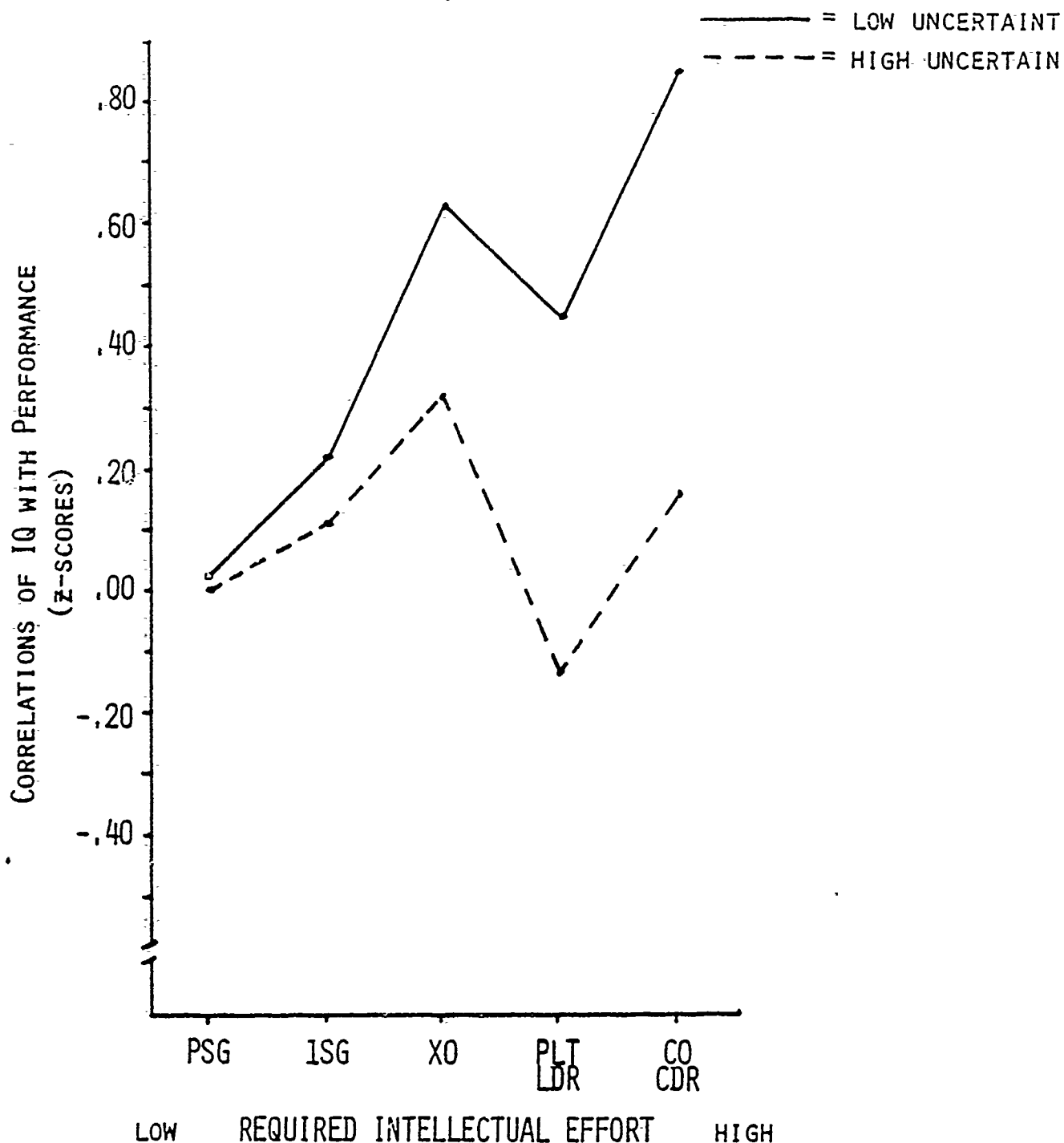
<u>INTELLECTUAL DEMAND OF TASK</u>	<u>IQ x PERFORMANCE</u>		<u>EXPER. x PERFORMANCE</u>	
	<u>LOW BOSS STRESS</u>	<u>HIGH BOSS STRESS</u>	<u>LOW BOSS STRESS</u>	<u>HIGH BOSS STRESS</u>
<u>HIGH</u>				
POLICY ADVISING	.27	-.46**	.06	.43**
PROJ. ENGINEERING	.38	-.16	.30	.13
EVAL. FIELD UNITS	.20	-.36	-.07	.60**
DECISION MAKING	.11	-.47#	.01	.47**
ADMIN. SUPPORT	.35*	-.47**	.06	.40**
MEAN	.26	-.39	.07	.41
<u>LOW</u>				
ATTENDING STAFF MTG.	.28	-.28	-.06	.53**
SUPERV. SUBORDINATES	.07	.04	-.06	.41*
TRNG. FIELD UNITS	.11	-.17	-.09	.42**
PUBLIC REPRESENT.	.04	-.36	-.09	.55**
ROUTINE PAPERWORK	-.01	-.25	-.28	.44**
MEAN	.10	-.21	-.12	.47
<u>RANK ORDER</u>				
CORRELATIONS WITH INTELLECTUAL DEMAND	.75	-.31	.80	-.18

* P < .05

** P < .01

P < .10

FIGURE 4



CORRELATIONS BETWEEN INTELLECTUAL ABILITY AND PERFORMANCE OF LEADERS IN DIFFERENT ARMY INFANTRY JOBS UNDER CONDITIONS OF HIGH OR LOW STRESS DUE TO UNCERTAINTY WHEN JOBS ARE ORDERED ON THE BASIS OF INTELLECTUAL EFFORT REQUIRED OF THE LEADER.

PSG = PLATOON SERGEANT

ISG = FIRST SERGEANT

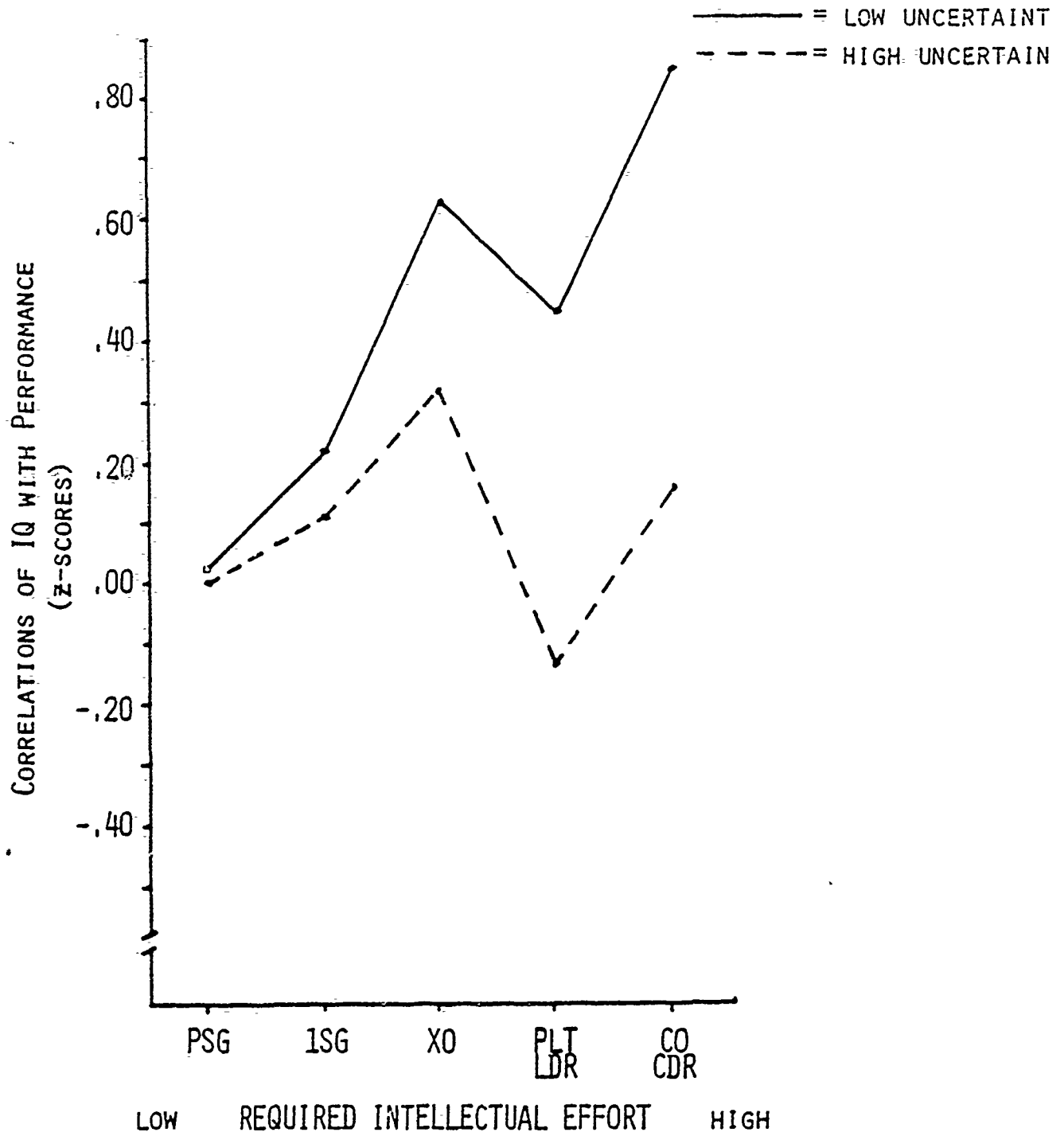
XO = COMPANY EXECUTIVE OFFICER

PLT LDR = PLATOON LEADER

CO CDR = COMPANY COMMANDER

Figure 5. Correlations between leader intelligence scores and performance ratings under low and high stress conditions, with jobs categorized on the basis of intellectual demand

FIGURE 4



CORRELATIONS BETWEEN INTELLECTUAL ABILITY AND PERFORMANCE OF LEADERS IN DIFFERENT ARMY INFANTRY JOBS UNDER CONDITIONS OF HIGH OR LOW STRESS DUE TO UNCERTAINTY WHEN JOBS ARE ORDERED ON THE BASIS OF INTELLECTUAL EFFORT REQUIRED OF THE LEADER.

PSG = PLATOON SERGEANT

1SG = FIRST SERGEANT

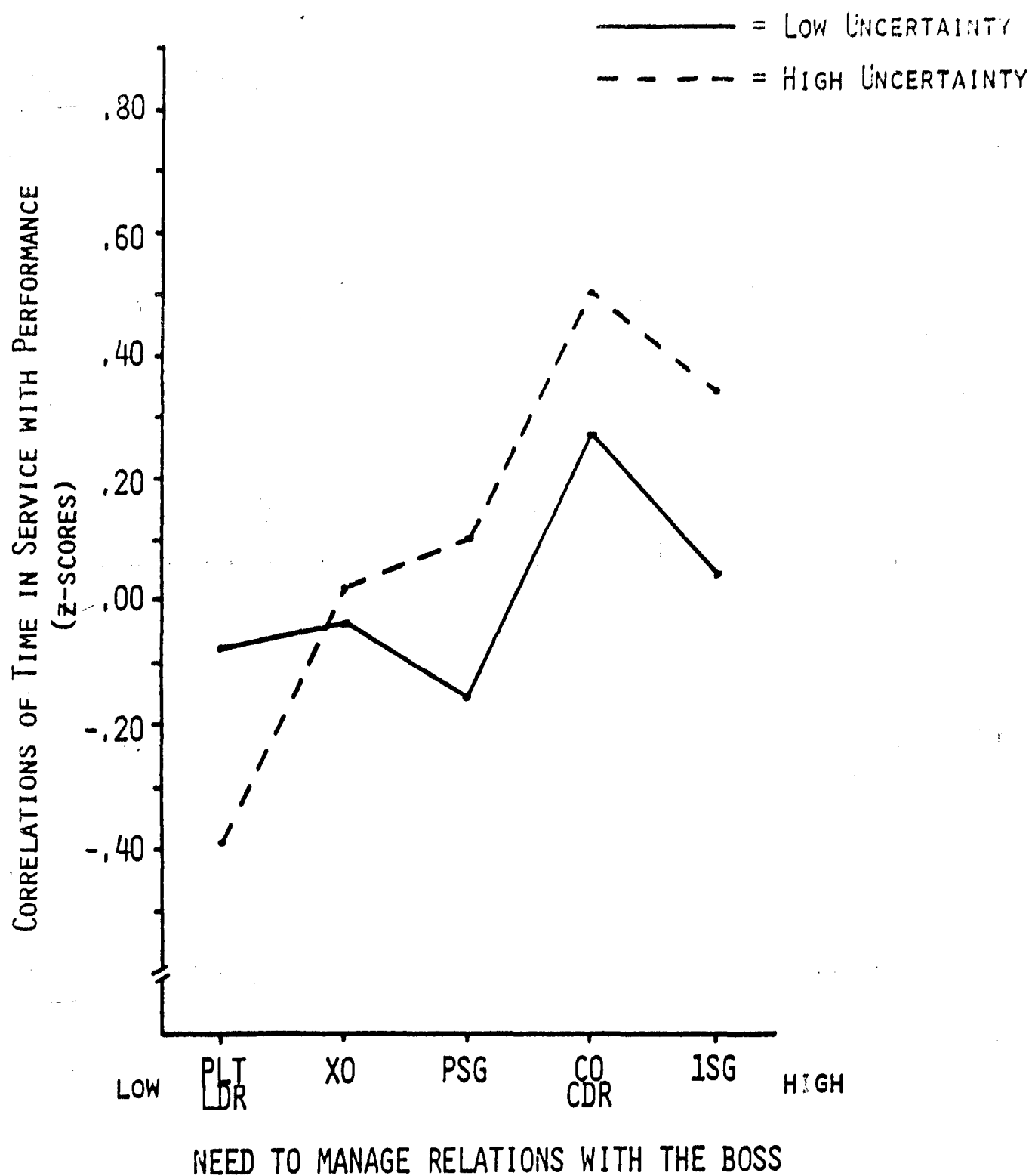
XO = COMPANY EXECUTIVE OFFICER

PLT LDR = PLATOON LEADER

CO CDR = COMPANY COMMANDER

Figure 5. Correlations between leader intelligence scores and performance ratings under low and high stress conditions, with jobs categorized on the basis of intellectual demand

FIGURE 5



PSG = Platoon Sergeant

PLT LDR = Platoon Leaders

1SG = First Sergeant

CO CDR = Company Commander

XO = Company Executive Officer

Figure 7. Correlations between leader experience (Time in Service) and performance under low and high stress conditions, with jobs categorized on the basis of required ability to deal with boss.

Symposium Luncheon Address

Sharon B. Lord, Deputy Assistant Secretary for Equal Opportunity
and Safety Policy, Department of Defense

I'm honored to be here to address the participants in the eighth "Psychology in the Department of Defense Symposium." My experience of consulting with components of DoD through the years has brought me into contact with numerous psychologists and counselors who were and are serving their country well by contributing their behavioral science expertise to the military. In a newspaper interview recently, I was asked by the reporter whether defense provided attractive career opportunities for psychologists. When I answered affirmatively, I pointed out that I felt defense needed us. Since I have assumed my current position I have been troubled by the lack of data in many areas where we are seeking answers and direction for vital human resource management questions.

I have been perturbed by the naivete of the paradigms used to approach or frame key research questions/problems. I have been frustrated by the lack of familiarity with important data and principles which I believe would dramatically enhance the productivity and quality of our force. (An example would be the way we systematically contribute to the stress of personnel.) On the other hand, I have been delighted at the receptivity and actions of many of the top level military and civilian leaders when new information or proposals are presented to them. And so I am encouraged and heartened in my belief that behavioral scientists have much to offer defense as we strive toward managing productively all of the human resources available to DoD, military and civilian.

As we have become more sophisticated in our understanding of the challenges we face in effectively managing an integrated force and by that I mean gender integration, racial integration and ethnic integration, it has become obvious that it is not sufficient to examine hispanic issues, women's issues, or any specific group's issues in isolation - separate from the personnel management of the total force.

Example - Army Study on Women.

1. Twenty-nine women's issues.
2. Major personnel policy review.
3. Four gender issues - for example, soldier/soldier marriages - as a women's issue.
4. One female unique issue - pregnancy.

The army will benefit - but negativity focused on women unnecessarily. Taking a human resource management approach would mean we would focus on the challenges and the positive aspects related to management of a gender integrated force.

In order to look at the human resource management challenge systematically, let's take a look at the human goals statement of the Department of Defense to see how well we are accomplishing these goals and also to see where the contributions of psychologists are much needed.

The document proclaims "The defense of the nation requires a well-trained force, military and civilian, regular and reserve.....We must increase the attractiveness of a career in defense so that the service member and the civilian employee will feel the highest pride in themselves and their work, in the uniform and the military profession." It addresses specific goals to achieve in each part of a service member's career from accession to transition or return to the society at large. Before I present my thoughts about those specific goals, let me point out that defense, in striving to attain the goals, runs into many kinds of challenges and questions that are similar to those raised in the larger domain of American society. These issues range from economic issues to the less structured, harder-to-define "issues" of equal opportunity. Obviously, societal values, successes and problems directly affect the success we have in achieving the goals outlined in the human goals proclamation. In short, we face human resource management challenges similar in many ways to those faced in the larger American society we are sworn to protect.

The human goals proclamation is the cornerstone of the philosophy under which defense operates. We are dedicated to the fact that to accomplish the mission, we must create an environment free of discrimination, either personal or institutional. Everything we do in acquiring, utilizing, developing, and managing human resources must be based on that philosophy. That's why Secretary Weinberger, the service secretaries, and the Joint Chiefs participated in a highly visible ceremony at the outset of this administration - to reaffirm this charter. So, as I review the specific goals and their implications, I will be surfacing human resource management issues and questions - for your consideration. The first of these goals is to attract to the defense service people with ability, dedication, and capacity for growth. Are we doing this with regard to the society's vast human resource pool in its entirety? I believe we must examine more carefully to see if we are "targeting" our recruitment efforts at segments of the population because of a specific definition of "ability, dedication, and capacity" without taking into consideration how our efforts or strategies have considered the issues of race, national origin, ethnicity, religion, and sex. We could very well be missing countless people who can and would willingly contribute to the defense of the nation. To what extent does our use of the media unintentionally send a message that we are looking for a specific type of person - others need not call. Do we portray military service situations in such a way that service to the country appeals to only a certain segment of our society? To what extent does language used, occupations portrayed, role models presented, and, indeed, skill levels being presented affect our attainment of this goal? Do our recruitment efforts "network" into the barrios, ghettos, college dormitories, sporting events, and high schools to the extent that no one is denied the opportunity to try to demonstrate that they are able, willing and ready to serve the nation?

The second goal, "To provide opportunity for everyone, military and civilian, to rise to as high a level of responsibility as possible, dependent only on individual talent and diligence," Can be translated as, "How well do we utilize our human resources - from training to release from active service?" It is very easy for some to let only the field commanders worry about the question. I believe, however, that we who support the field commander must continue to examine larger questions. One such question is, "Do we pay enough time and attention to the maintenance of our human resources - as much as we do the maintenance of weapons?" This question can be subdivided into several dimensions:

For example: Training, I believe much research is needed to determine whether we effectively teach and deliver the concepts and information we believe to be important "roots" versus mobility (Here I am curious about the effects of constant changes of station on the service members and their dependents), organizational versus personal needs - (How do they intersect), and finally who defines the standard of "Talent and diligence?" Looking at maintaining our human resources from the angle of "Maintaining the psychological health, physical productivity and freedom from stress or dis-ease necessary for optimum productivity. I would argue, as I mentioned earlier, that we have devoted far too little effort and research to maintaining standards for high performance human technology. We maintain our hardware in a more sophisticated fashion than our human resources.

The third goal and the one I would like to highlight is "To make the military and civilian service in the Department of Defense a model of equal opportunity for all regardless of race, color, sex, religion or national origin." In response to this goal, defense has initiated numerous people-oriented programs, refined affirmative actions programs, raised the cultural awareness of its members through such things as changes in policy and procedures, the establishment of the Defense Equal Opportunity Management Institute, and redefining for field commanders their role as catalysts toward achieving this goal. Yet, we have not been fully successful in this endeavor and there remain countless questions to be answered. We are under constant scrutiny by society in this effort. Some may believe that this goal is out of reach or inappropriate. I believe that we need to translate what we have called equal opportunity into leadership concepts and human resource management concepts. A workplace free of undue harassment affects productivity positively. An environment free of racial slurs and bias contributes to cohesion. Equal access to promotion affects retention.

When the force lives and works in an environment or a field that (as the human goals statement so aptly states) is based on the principle that the individual has ultimate dignity and worth - we aren't just being fair and nice - I would argue that we are creating a force field where human productivity is maximized. A powerful force field necessary to the nation's security. Secretary Weinberger, through Dr. Korb, the Assistant Secretary for Manpower, Reserve Affairs, and Logistics (Dr. Korb) has been extremely active in seeking the attainment of these goals. We have implemented or are seeking changes in policy and direction in many areas I have mentioned. An example of one of these efforts lies in the ever-expanding mission of the Defense Equal Opportunity Management Institute (EOMI) located at Patrick Air Force Base, Florida. They are in the business of helping defense deal with the challenges of managing a fully integrated force - or as I put it - of creating a force field free of arbitrary barriers to productivity - so that we are maximizing each and every human resource.

As you know, the institute exists primarily to train people of all services for assignments as equal opportunity staff members. It provides our commanders with persons who are (1) sensitive to the issues affecting these human goals, (2) adept in identifying the commander's day-to-day concerns about the people in the unit, and (3) astute in finding ways and means of insuring "unit cohesion," a term I borrow from the army. EOMI's effectiveness has increased through the years. Its reputation among other governmental agencies both locally and as far away as the London police department demonstrates the importance of its support role. In fact, I have just recently approved the institute's developing an equal

opportunity training plan for NASA. Major commanders of the services have responded to the institute's support role by asking the institute to update their own trainers (Staff officers' course and the training of Navy EEO personnel). In the academic community, EOMI's participation in gatherings such as this has raised its professional credibility. However, the issues are too fluid and numerous for any one agency to handle. So let's examine another facet of the human resource management challenge-research.

Before I point out some of the major issues which I believe need to be analyzed, I want to explain the researchers' role in supporting the commander. Recall for a moment that the ultimate mission of the Department of Defense is to provide for the security of the nation. I have already addressed some of the goals we must reach to assist in that endeavor. The researchers' role in the accomplishment of that mission is to provide the Department of Defense and field commanders with information which identify problems and provide solutions. As I stated in an earlier speech, the management of human resources is a complex process, that cannot be left to chance, the well-intended, or the not-so-well intended. It is a process that could be managed as effectively as the development and introduction of new weapons, if the proper variables and information are considered. We must be able to manage the contribution of all our resources, men and women, minority and majority, protestant, catholic, jew, or hindu in a way that aids them in moving beyond biases that they may have toward one another. Researchers must play an important role.

The research challenges which I am about to deliver are focused primarily on the human goals which, in part, states we must strive for greater utilization of our human resources. Let's look at key issues which affect the service member from training, through tenure, to transition back to civilian life.

In Training Research:

I believe that the service member's training experience sets the tone and direction for his or her career. Thus, I am suggesting some areas for further exploration. Such questions need answering as:

- a. Does the delivery of military training contain institutional discrimination in any way, shape, or form?
- b. Is military training constructed with equal opportunity inhibitors? How do language, gender, race, etc., impact on the design and delivery of training at various stages of a person's career?
- c. Does there exist in the training environment the tendency to focus on the victim of discrimination? Do we (or should we) change training objectives to accommodate selected groups of people, i.e., lower to assure success or raise to insure failure?
- d. Do we unconsciously deliver training based on individual values, beliefs, and attitudes of the trainer or curriculum developer which impedes education of those with different values?

In Tenure Research

As the service members move to the next point in the life cycle of their tenure,

they experience many things that impact on their performance of duty. Unfortunately, some of those experiences are inhibiting their decisions to make military service a career. So, in the area of maintenance of the human resource, we need to:

- a. Further explore what is inhibiting job satisfaction in the military.
- b. Obtain data that accurately tracks the changing nature of the military family and how family situations with military life affect retention of military members.
- c. Determine how we are operationally defining merit, talent, and success. How do we go about insuring that our definition of these terms are understood by all?
- d. Define the exact impact that frequent changes of station have on our service members and their families. In other words, are they necessary for mission accomplishment or are they just "traditional?" I believe the Chief of Staff of the Army, General Meyer is beginning to have second thoughts about frequent individual moves as reflected by his decision to test the regimental system again.
- e. Find programs and procedures which tend to match individual goals with the needs of the military. I am looking for a means of convincing our people that serving in the military is a step toward - not away from - their own hopes, dreams, and ambitions.
- f. Open a direct, clear channel of communication with field commanders to let them know what the situation is, not what it has been or what it is going to be. What I am asking you to look at here is finding a way to apprise our commanders of current issues - and answers - which impact on their human resource. We do just that when we keep them informed on weapons technology financial parameters, and even strategic developments. I believe there must be a way. We have learned already that it does no good to care for munitions, equipment, and spare parts unless we pay equal attention to the commanders' people and the quality of their life in the military. Research to promote psychological and physical wellness in our force.

In Transition Research:

When we examine how we are doing to insure that we return to society people who know they can, and will, contribute, there remain some areas which need research:

- a. What caused the individual to leave the service? Was it system-generated or was the departure caused by outside factors? Was it because we had not met the individual's expectations or, indeed, did we set up unreasonable organizational expectations?
- b. Do we have a feedback system to check on the impact we may have had on former service members? To what extent did military service prepare the person for successful participation in our society?
- c. We need to define the commanders' role in the transition process.

What part, if any, does the commander have in insuring the departing servicemember has been prepared properly for the transition? Have we, in fact, provided commanders the skill and resources to fulfill that task?

d. Apart from generally knowing the skills of the people we are sending back to society, do we know exactly what skills are being discarded? Let me clarify that. We may know by computer printout that 2,000 people carrying the job code infantry are leaving the service next month. Is there a way of inventorying the "other" skills which those people are taking with them? Are we letting people leave without having reflected on their attitudes prior to departure?

Those are just a few of our challenges. I hope that while you're here and once you have returned to your respective positions, you will find the challenges stimulating enough to spur you into finding answers to the issues I have presented.

Defense of the nation.....A tough job. All of us are involved. All of us must strive to use our human resources toward that mission in the most efficient and effective manner possible. In doing so, we cannot afford to forget or disregard the individual's worth and dignity. More importantly, it is wrong to discount any member of our department - or our society for any reason. We cannot waste a single moment in achieving full implementation of the human goals charter. A productive fully integrated, powerful force field. Our nation can settle for no less.

Structures of Memory for Critical Flight Information

Roger W. Schvaneveldt and Timothy E. Goldsmith
Department of Psychology
New Mexico State University
Las Cruces, NM 88003

and

Francis T. Durso
Univeristy of Oklahoma

Abstract

Critical flight-related concepts were subjected to multidimensional scaling procedures. These structural representations are shown to possess information that can discriminate members of different groups of flying personnel. The concepts of experts and novices were analyzed to determine how the organization of concepts for novices compares to the organization found for experts. Multidimensional scaling solutions allow for better discrimination than direct rating data. It is concluded that these methods can be used successfully to analyze the conceptual structure necessary to perform a complex task. Further, the structures themselves may prove to be useful in training.

Introduction

Research in cognitive psychology has shown that retrieval of information from memory is affected by the organization of that information in memory (e.g., Meyer & Schvaneveldt, 1976; Schvaneveldt & Meyer, 1973). More rapid retrieval results from a more efficient and economical storage structure. The organization of information in memory can have a vital impact on flying performance. An understanding of the memorial organization of critical flight information can be extremely useful to training program designers and evaluators as well as instructors. Knowledge about the conceptual structures held by expert fighter pilots can be used to tailor training systems to provide students the conceptual framework needed to operate fighter aircraft effectively. Such knowledge may also provide a useful selection tool by allowing instructors to determine which individuals have mastered the concepts required for success in a particular training program.

The objective of our studies is to define structures of memory for critical flight information in Air Force fighter pilots. We used various structural analyses with groups of flying personnel differing in the extent and nature of flying experience. In this paper, we will discuss our work using multidimensional spatial representations for one set of concepts from split plane maneuvers. Multidimensional spatial representations (Shepard, 1963; Kruskal, 1977) show the location of each concept in a multidimensional psychological space where the Euclidean distance between concepts represents psychological proximity. The validity of these representations is assessed by various criteria, including consistency within and between individuals and the development of the representations with training and experience.

MULTIDIMENSIONAL ANALYSIS

One set of concepts was presented on the console of a Terak 8510/a microcomputer to four groups of officers: Air National Guard pilots (GP), Fighter Lead-In Instructor pilots (IP), recent Undergraduate Pilot Training graduates (UP), and Instructor Weapons Systems Officers (IW). Multidimensional scaling (MDS) was used to produce structural descriptions of the concepts. The concepts selected from split-plane maneuvers are shown in Table 1. The set of concepts has an associated scenario to define the context in which to consider the concepts. The 30 basic concepts were presented in a rating task. The subject groups are summarized in Table 2.

Each subject rated the similarity or relatedness between the members of each pair of concepts. Each of the 435 pairs was rated on a 10 point rating scale. The obtained similarity measures were transformed into measures of psychological distance by subtracting the ratings from the maximum possible rating. The rating task provided the data base for scaling. Since some of the subjects tested were in UPT training, a familiarity rating task was used to determine whether the UPTs were familiar with the terms. Overall, at least some familiarity was claimed for 63% of the concepts.

Dimensions in Multidimensional Space. Some of the major dimensions have been identified for the MDS solutions for IPs. The split-plane concepts have two dimensions associated with temporal factors and one dimension that distinguishes particular maneuvers. The first temporal dimension identifies the general time dimension within a scenario leading to split-plane maneuvers. The second temporal dimension represents the ordering of concepts in a standard training sequence which instructors and students frequently follow in practice maneuvers. The third dimension represents a contrast between lead pursuit and lag pursuit with LAG PURSUIT and the associated maneuvers at one end of the dimension and LEAD PURSUIT and LOW YO YO at the other extreme. Temporal order is apparently important in the organization of these concepts for pilots. Dimensions based on temporal order show up quite consistently.

PATTERN RECOGNITION ANALYSIS OF CONCEPTUAL STRUCTURES

The objective in this phase of the project is to develop methods for classifying an individual as a member of a particular group based on conceptual structure. In addition to classification, the analysis also provides information about the degree to which each individual is associated with each group. The analysis to be described applies the principles and techniques of pattern recognition (Nilsson, 1965).

Pattern recognition is an area of artificial intelligence (AI) that is generally concerned with deciding whether an unknown object is a member of a particular class of objects. AI applications often involve computer identification of visual objects, although numerous other uses exist. Such a method allows for abstract as well as physical objects to be analyzed. In addition to categorizing objects as members of a particular class, pattern recognition principles also supply information about class and individual differences.

Two types of patterns were formed for each individual tested. One pattern was generated from the conceptual structures derived from a multidimensional scaling (MDS) analysis. The MDS pattern was created by the set of distances

Table 1. SCENARIO: SPLIT-PLANE MANEUVERS

<u>Assumptions</u>		
OFFENSIVE	AGGRESSIVE	TALLY HO
SINGLE BANDIT	KILL	EQUAL OR SIMILAR AIRCRAFT
COMMIT	ENGAGED	IR MISSILE PARAMETERS
DEFENSIVE TURN (TARGET DENIES MISSILE)		
<u>Basic Concepts</u>		
LOW YO YO	HIGH YO YO	QUARTER PLANE
LAG ROLL	BARREL ROLL	OVERTAKE
GUNS	AIRSPEED	ANGLE OFF
G LOADING	CUTOFF	RELATIVE ENERGY
6 O'CLOCK	SMASH	POWER SETTING
SWITCHOLOGY	ACCELERATION	RADIAL G
HEAT	SNAPSHOT	VERTICAL MANEUVERING
3-9 LINE	EXTENSION	WEAPONS PARAMETERS
LAG PURSUIT	LIFT VECTOR	CORNER VELOCITY
ASPECT ANGLE	PURE PURSUIT	LEAD PURSUIT
<u>Other Related Concepts</u>		
THREAT	DISENGAGE	BINGO
JOKER	SADDLED-UP	LOW PK
HIGH PK	BURIED NOSE	REVERSAL
SHOOTER	LOAD	PADLOCK
RTB	EGG	OP TURN
MAX TURN	TRAPPED NOSE	LINE OF SIGHT RATE
NOSE COUNTER	KNOCK-IT-OFF	OVER PULL
UNDER PULL	SEPARATE	SUN
CLOUDS		

Table 2. Summary of Subject Groups

Group	n	Description	Location
IP	7	Instructor Pilots	Holloman AFB
GP	9	Air National Guard Pilots	Buckley ANGB
UP	17	Undergraduate Pilots (UPTs)	Williams AFB
IW	4	Instructor Weapons Systems Officers	Holloman & Williams

between concepts in the MDS solution. This allowed the pattern to preserve the structural properties inherent in the MDS solution. A second pattern was constructed of the set of similarity ratings for all pairs of concepts. Both methods resulted in patterns with 435 features.

The first analysis consisted of applying a minimum-distance classifier to all pairs of groups. With this procedure, a prototype point representing the central tendency of a group of patterns is constructed for each group. The prototype point of a group is simply the average of the feature values of all patterns belonging to the group. A minimum-distance classifier computes the distances between each pattern to be categorized and each prototype point and places the pattern into that group associated with the nearest prototype.

In the case of two groups, the decision surface separating the patterns is the perpendicular bisector of a line connecting the two prototype points. This approach works well when the patterns of each group cluster tightly around their respective prototype points, and the group clusters are well separated. Prototype points for all groups and decision surfaces for separating all pairs of groups were computed. In each application of the minimum-distance classifier, all members of the two groups were used. This provided information about both group and individual differences. The distances from each individual to a decision surface and from each individual to the group prototypes can be determined as well as the distances between group prototypes.

The second analysis involved computing a decision surface that separated members of training sets consisting of a limited number of members from two groups and then applying the decision surface to the remaining members of the groups. Decision surfaces were computed with a training algorithm if a minimum-distance classifier based on the members of the training sets did not separate the training sets. The training procedure alters the surface between two groups by successive adjustments to a weight vector which in effect changes the orientation and position of the decision surface. In the case of two linearly separable groups, a weight vector exists that will produce a discriminant function that returns a positive value for all patterns from the first group and a negative value for all patterns from the second group. Weight vectors were initialized to the weights produced by a minimum-distance classification of the individuals in the limited training set. This procedure kept the final decision surface as close to the minimum-distance decision surface as possible. For each pair of groups, a training set of a particular size was randomly chosen, and a decision surface was computed to separate the members of the training set into their respective classes. The resulting discriminant function was then applied to the individuals who were not included in the training set. This procedure was repeated 100 times for each training set size. The training sets consisted of equal numbers of individuals from each group. The whole procedure was iterated with successively larger training sets until the size of the smaller group was reached.

The major finding was that pattern recognition techniques can be used to discriminate classes of flying personnel based on their conceptual structures of critical flight information. Also significant was the result that patterns represented by distances in an MDS solution produced better group separation than patterns based on rating scores.

A minimum-distance classifier applied to each pair of groups resulted in well separated groups with only a few erroneous classifications. When the minimum-distance classifier was applied to rating data, a small number of misclassifications occurred. For the split-plane maneuvers, all of the misclassifications occurred between IPs and UPTs. The two IPs who were classified as GPs under the ratings were on the GP side of the decision surface separating IPs and GPs. This means that their conceptual structures resembled more closely those of GPs than IPs. Four of the UPTs have conceptual structures that resemble more closely a GP structure than a UPT structure. An important point is that these misclassifications occur even though the individuals misclassified contributed to defining their group prototype.

The number of misclassifications occurring between two groups reflects both the distance between the groups in the pattern space and also the tightness with which individuals cluster around their respective prototypes. IPs and UPTs appear to be very distinct classes since no IPs were classified as UPTs and no UPTs were classified as IPs for either maneuver. Considering the simplicity of the decision surfaces produced by the minimum-distance approach, the overall results indicate quite distinct classes of individuals.

Using the MDS distances, we find perfect separation of all classes. Apparently the structural information supplied by the MDS procedure maximizes the differences between classes. This finding helps to validate the claim that the MDS technique extracts important structural information from similarity ratings.

The second analysis involved generating a discriminant function on the basis of a limited training set from two classes and then using the function to place new and unknown members into one of the two classes. An example of the results of the classifications is shown in Table 3. The table gives the total number of individuals for which a classification was attempted, followed by the percentage of those correctly classified and the probability of randomly classifying at least this number. Since 100 different randomly chosen training sets were used for each training set size, the number of classifications attempted is always 100 times the number of remaining members in the two classes.

The classification of unknown members was quite successful. With only one member each from the IPs and GPs on which to base a decision surface, 798 out of 1400 remaining IPs and GPs were classified correctly. Performance was even better for MDS patterns where 949 of the 1400 cases were correctly classified. These results are more impressive if we consider that two of the seven IPs resemble GPs. In general, classification improves as the distance between classes in the pattern space increases. With only two members each from the IPs and UPTs, it is possible to classify correctly the remaining 20 members 95 percent of the time.

Discriminant functions based upon the MDS patterns resulted in better classification of unknown members compared to the rating patterns. The average percentage of correct classifications for the MDS patterns was significantly greater than the average for the ratings, $t(106)=5.95$, $p .001$. As mentioned previously, this superior performance can be attributed to the additional structural information supplied by MDS.

Table 3

Classification of Groups Members
on the Basis of Limited Training Sets

Split-Plane Maneuvers

Training		#Classified	Ratings		Distances in MDS	
Set	Size		%Correct	Prob*	%Correct	Prob*
IPs	GPs					
1	1	1400	57	p .001	68	p .001
2	2	1200	67	p .001	77	p .001
3	3	1000	67	p .001	82	p .001
4	4	800	68	p .001	84	p .001
5	5	600	62	p .001	87	p .001
6	6	400	52	p .05	87	p .001
7	7	200	51	p .05	84	p .001
Total		5600	63	p .001	79	p .001

Training		#Classified	Ratings		Distances in MDS	
Set	Size		%Correct	Prob*	%Correct	Prob*
IPs	IWSOs					
1	1	900	53	p .05	56	p .001
2	2	700	57	p .001	60	p .001
3	3	500	61	p .001	63	p .001
4	4	300	71	p .001	67	p .001
Total		2400	58	p .001	60	p .001

Training		#Classified	Ratings		Distances in MDS	
Set	Size		%Correct	Prob*	%Correct	Prob*
IPs	UPTs					
1	1	2200	65	p .001	79	p .001
2	2	2000	77	p .001	95	p .001
3	3	1800	78	p .001	98	p .001
4	4	1600	83	p .001	100	p .001
5	5	1400	89	p .001	100	p .001
6	6	1200	91	p .001	100	p .001
7	7	1000	96	p .001	100	p .001
Total		11200	80	p .001	94	p .001

* Probability of randomly classifying at least the number of individuals correctly classified.

Pattern recognition techniques appear to be sensitive enough to detect subtle differences between both groups and individuals. In addition, these differences often seem to have "real-world" significance. Many of the findings may have relevance to selection and training.

CONCLUSIONS AND RECOMMENDATIONS

This project has produced several interesting and potentially useful findings. In the domain of critical flight-related concepts, we have shown that systematic methods can yield valid and reliable descriptions of the structure of these concepts. Further, these structures can be used to identify individuals as members of groups with different training and experience. The structures should be of use in designing training programs for students attempting to acquire these conceptual structures. In addition, the representations themselves may prove to be useful as training aids.

Our work in classifying individuals based on their conceptual structure suggests further work in attempting to predict the success of future fighter pilots based on the conceptual structures students demonstrate early in training. It may be necessary to study the structures associated with a different set of concepts than those used in the present investigation. For example, some concepts relating to attitude and motivation could be included along with concepts relating to the operation of aircraft. The classification techniques we have developed appear to be very sensitive to differences in cognitive structures, and they may well provide some predictive power for organizing the training program to produce maximum benefit for those who are likely to benefit the most from fighter training.

REFERENCES

- Kruskal, J. B. Multidimensional scaling and other methods for discovering structure. In Enslein, Ralston, and Wilf (Eds.), Statistical Methods for Digital Computers. New York: Wiley, 1977.
- Meyer, D. E., & Schvaneveldt, R. W. Meaning, memory structure and mental processes. Science, 1976, 192, 27-33.
- Nilsson, N. Learning Machines: Foundations of Trainable Pattern-Classification Systems. New York: McGraw-Hill, 1965.
- Schvaneveldt, R. W., & Meyer, D. E. Retrieval and comparison processes in semantic memory. In S. Kornblum (Ed.), Attention and Performance IV. New York: Academic Press, 1973.
- Shepard, R. N. Analysis of proximities as a technique for the study of information processing in man. Human Factors, 1963, 5, 33-48.

A Video Game as a Covariate for Carrier Landing Research

Gavan Lintern and Robert S. Kennedy

Canyon Research Group, Inc.
1040 Woodcock Road, Suite 227
Orlando, Florida 32803

Abstract

Statistical power is an important although neglected concept in experimental design. Briefly, an experiment should have adequate power so that there is a reasonable chance of demonstrating statistical significance in the presence of real effects, and of being able to assert with some confidence, that there is no effect if one is not demonstrated. Unexplained experimental variance is particularly damaging to power. One means of reducing that variance is to assess subject aptitude for the task to thereby account for some of the between-subject variance through an analysis of covariance.

This paper describes exploration of a commercially available video game as a covariate for training experiments with a full scale carrier landing simulation. The video game was selected because prior research had shown a high test-retest reliability for practiced subjects, and also that terminal performance correlated well with terminal performance on a compensatory tracking task. The data reported here examine the test-retest reliabilities of performance with the video task. The relationships between the video task and one experimental dependent measure, and the test-retest reliability of that dependent measure are also examined.

Introduction

Power analyses and their value have been discussed by Cohen (1977). Briefly, an experiment should have adequate statistical power so that there is a reasonable chance of demonstrating statistical reliability in the presence of real effects. In a review of training effectiveness of visual and motion simulation, Waag (1981) has been constrained to tentative and partial concluding statements about issues that have been the focus of extensive flight simulation research. The fact that the final section of his report is substantially devoted to methodological difficulties with transfer-of-training research reflects his view of the cause of this unsatisfactory situation. There can be little doubt for the experienced aviation psychologist that his view is valid. One of the problems highlighted in his discussion was that of inadequate statistical power.

Unexplained experimental variance is particularly damaging to power. Research experience with the Visual Technology Research Simulator (VTRS) has indicated that subject differences can account for 25% or more of the

variance (Westra, Simon, Collyer, Chambers, & Nelson, 1981). One means of explaining that variance is to assess subject aptitude for the task to thereby improve the power of the experiment with an analysis of covariance. Unfortunately it has been difficult to identify a good covariate for flight tasks.

A good covariate should correlate well with the experimental dependent measure. Terminal performance on Air Combat Maneuvering (an ATARI video game) has been shown to correlate well with terminal performance on a compensatory tracking task (Kennedy, Bittner, & Jones, 1980). As compensatory tracking is an integral part of many flight tasks, Air Combat Maneuvering (ACM) was selected for further study.

Another positive feature of ACM is its high test-retest reliability (Jones, Kennedy, & Bittner, 1981). Note however that the high reliability was demonstrated only for practiced performance and that the reliability of unpracticed performance was substantially lower. The two-process theory of motor learning (Jones, 1970) states that the rate of acquisition and subsequent terminal performance are two separate, and one may infer, unrelated processes. Thus practiced performance may be a good covariate where unpracticed performance is not. This is more likely to be true where, as in the case of flight training research, the covariate is intended to account for subject variation in terminal performance on the criterion task.

Three sets of data are reported that bear on the potential value of ACM as a covariate for carrier-landing research. The first set is used to check the findings of Jones, *et al.* (1981) that the test-retest reliability of ACM is high and stable after approximately three hours (60 trials) of practice. The other two are from training experiments for carrier landings; one involving military pilots and the other flight-naive college students and graduates. These data are used to look for changes in test-retest reliability of ACM with practice, to examine the test-retest reliability of the metric for carrier-landing performance, and to ascertain the correlation between ACM and carrier-landing performance.

Method

Experiment 1. Nineteen flight-naive college students with 20/20 corrected vision were tested with a two-dimensional pursuit tracking task known as Air Combat Maneuvering (ACM), commercially available as an ATARI video game (Combat Type CX2601, Game #24, difficulty 'b', right controller). Details and procedures were as described by Jones, *et al.* (1981) except that sixty trials were administered in six blocks of ten trials. The task was scored on the number of hits per 2.25 minutes trial.

Experiment 2. Thirty-three military pilots were taught carrier landings in the Visual Technology Research Simulator (VTRS): a high-fidelity simulator with an instrumented T-2C Navy jet trainer cockpit, a six degree-of-freedom synergistic motion platform, a wide angle visual system that projects a computer-generated image, and an Experimenter/Operator Control Station (Collyer and Chambers, 1978). None of the pilots had attempted carrier landings in an aircraft or simulator; nor had they flown approaches with

the Fresnel Lens Optical Landing System, which is the primary glideslope guidance system for a carrier approach. Their instruction in the VTRS consisted of forty training trials with between-subject variations in equipment configurations and instructional methods.

Following their training trials all pilots were transferred to a criterion VTRS condition for sixteen trials. The criterion condition included a 160° lateral by 80° vertical field of view and a day carrier scene, and was flown with moderate turbulence and with platform motion. In that condition, pilots started on the downwind leg (heading opposite to the carrier heading), at 600 feet altitude, 5700 feet from the carrier, flying straight and level, and in the landing configuration. The pilots were expected to fly past the carrier and then to turn through 190.5° to intercept the 3.5° approach glideslope when they were behind the carrier. From there they were to fly to the carrier landing deck by following the glideslope. Carrier-landing performance was assessed over the final 3000 feet of the approach by scoring the proportion of time spent within a tolerance band extending 0.3° above and below the glideslope.

The pilots were also given thirty to fifty ACM trials in ten-trial blocks which were interspersed between their sessions in the simulator.

Experiment 3. Six of the college students from Experiment 1 were selected, on the basis of their availability, for carrier-landing training in the VTRS. This group was supplemented with two flight-naïve college graduates who were tested with ACM as described for Experiment 1.

These subjects were taught carrier landings in the VTRS (three blocks of 15 trials) with a day carrier scene, with no turbulence, and with the motion system inactive. Each carrier approach was started 9000 feet behind the carrier, on the 3.5° glideslope and extended centerline of the landing deck, and in the landing configuration. Deviations from the glideslope were sampled at 30 Hz throughout the approach and used to calculate RMS error.

Analyses

Four-trial means were used to derive individual scores for glideslope tracking transfer performance in Experiment 2, and five-trial means were used for all other measures. The final four- or five-trial block of simulator trials was used to score glideslope tracking performance. (For Experiment 2, inspection of the data indicated that differences due to training variations had washed out early in transfer.) The first ACM five-trial block was used to assess unpracticed ACM, and the last (trials 26 to 30 for Experiment 2, and trials 56 to 60 for Experiment 3) to assess practiced ACM.

The correlations between glideslope tracking and ACM were corrected for the attenuation due to error variance in each (Guilford, 1954, p 400). Test-retest reliabilities were estimated from correlations with adjacent five-trial blocks. For Experiment 2, one block of trials was used to estimate the association between the two measures, or one adjacent block to estimate their reliabilities. The small number of subjects in Experiment 3 prompted the use of two adjacent blocks to stabilize those data for these estimates. The resulting correlations were averaged to estimate the required cross correlations and reliabilities.

Results

The magnitude of the five-trial means for ACM increased throughout training. The differences between the college students of Experiment 1 and the military pilots of Experiment 2 were small and were not statistically reliable. However both of the groups showed an advantage of approximately 3.5 hits per trial over the subjects tested by Jones *et al.* (1981). This advantage was maintained over 50 trials, and remained statistically reliable throughout ($p < .01$).

Experiment 1. Table 1 shows the intercorrelation matrix of ACM scores. Correlations are high, with a tendency to be higher and more consistent later in training.

Experiments 2 and 3. Table 2 shows correlations of simulator performance with practiced and unpracticed ACM. Also shown are estimated test-retest reliabilities of the dependent measures, and their cross correlation corrected for attenuation..

Correlations between glideslope tracking and ACM were modest, ranging from .35 to .48. However the low reliabilities of the glideslope tracking measure ($\approx .30$) were also modest, while the ACM reliabilities were high ($\approx .80$). After correcting for attenuation due to unreliability of the measures, correlations between the reliable components of glideslope tracking and ACM ranged from .69 to .95. There were no obvious differences between our military pilot sample and our college students/graduates sample. Nor were there any obvious differences in the corrected or the uncorrected cross correlations resulting from the use of unpracticed versus practiced ACM.

Discussion

The differences between our ACM data and those of Jones *et al.* (1981) were unexpected both in terms of the higher scores and the strength of the early correlations. It is not possible to ascertain whether there were basic aptitude differences between our college students and their Navy enlistees or whether increased experience with video games in the general population was responsible for the higher scores.

The Jones *et al.* (1981) data have led us to anticipate that practiced performance would provide a more reliable covariate than would unpracticed performance. However our data from both military pilots and college students/graduates indicate that initial ACM trials are as reliable as later trials.

The correlation of ACM with glideslope tracking was modest. However after correction for attenuation due to test unreliability, ACM was shown to account for 48% to 86% of the reliable variance in glideslope tracking. Thus the choice of ACM as a covariate seems to have been a good one, and the modest correlation between it and glideslope tracking appears to be due primarily to low reliability in the glideslope tracking measure. This indicates that statistical power in our research would benefit from development of a more reliable dependent measure.

In general, covariates have not been exploited in simulation research, even though individual differences typically account for the bulk of the variance. This may be due to their failure to live up to their theoretical

promise to improve statistical power. Unfortunately covariates for applied research are often chosen on the basis of intuitive appeal, and their measurement properties and the measurement properties of the dependent measures are seldom examined. As a result it is not known whether a covariate has failed because the assumption of an association between the covariate and the dependent variable is incorrect, or because of unreliable measurement of one or both of these. The implication of the data presented here is that prior exploration of potential covariates can establish which ones are likely to be useful and will identify the nature of their limitations.

Acknowledgment

The data for Experiment 2 were collected by Dr. Daniel P. Westra, as a part of a larger study. The data for Experiment 3 were collected by Daniel J. Sheppard, as a pretesting effort for another experiment. Brian Nelson, Karen Thomley and Daniel J. Sheppard otherwise assisted with data collection and analysis.

References

- Cohen, J. Statistical Power Analysis for the Behavioral Sciences, New York, Academic Press, 1977.
- Collyer, S.C., and Chambers, W.S. AWAVS, a research facility for defining flight trainer visual requirements. Proceedings of the Human Factors Society 22nd Annual Meeting, Detroit, Mich.: Human Factors Society, 1978.
- Guilford, J.P. Psychometric Methods (2nd ed.) New York: McGraw Hill, 1954.
- Jones, M.B. Rate and terminal processes in skill acquisition. American Journal of Psychology, 1970, 83, 222-236.
- Jones, M.B., Kennedy, R.S., and Bittner, A.C., Jr. A video game for performance testing. American Journal of Psychology, 1981, 94, 143-152.
- Kennedy, R.S., Bittner, A.C., Jr., and Jones, M.B. Video games and conventional tracking. Perceptual Motor Skills, 1980.
- Waag, W.L. Training effectiveness of visual and motion simulation. Brooks Air Force Base, TX: Air Force Human Resources Laboratory, AFHRL-TR-79-72, 1981.
- Westra, D.P., Simon, C.W., Collyer, S.C., Chambers, W.S., and Nelson, B. Investigation of simulator design features for the carrier landing task. Orlando, FL: Naval Training Equipment Center, Technical Report NAVTRA-EQUIPCEN 78-C-0060-7, 1981.

TABLE 1. INTERCORRELATIONS FOR FIVE-TRIAL MEANS
FOR SIXTY TRIALS ON AIR COMBAT
MANEUVERING (19 COLLEGE STUDENTS)

Mean of Trials	6-10	11-15	16-20	21-25	26-30	31-35	36-40	41-45	46-50	51-55	56-60
1-5	.89	.91	.90	.85	.86	.81	.84	.83	.78	.80	.83
6-10		.93	.94	.80	.85	.80	.84	.87	.87	.84	.87
11-15			.93	.84	.93	.88	.90	.90	.86	.88	.88
16-20				.87	.88	.88	.89	.88	.84	.86	.88
21-25					.93	.85	.94	.89	.86	.82	.85
26-30						.91	.96	.95	.86	.88	.87
31-35							.96	.92	.85	.93	.92
36-40								.95	.88	.91	.93
41-45									.91	.94	.94
46-50										.91	.92
51-55											.96

TABLE 2. INTERCORRELATIONS, ESTIMATED RELIABILITIES, AND CORRECTED
INTERCORRELATIONS FOR GLIDESLOPE TRACKING AND AIR COMBAT MANEUVERING.

	Experiment 2		Experiment 3	
	Unpracticed ACM	Practiced ACM	Unpracticed ACM	Practiced ACM
Glideslope by ACM Correlation	.48	.43	.35	.37
Glideslope Reliability		.32		.29
ACM Reliability	.81	.87	.89	.74
Glideslope by ACM Correlation Corrected for Attenuation	.95	.81	.69	.79
Reliable Common Variance	86%	66%	48%	62%

Standard Errors Associated
with Item Response Theory Parameters

Benjamin A. Fairbank, Jr., Ph.D.
5825 Callaghan Road, Suite 225
McFann-Gray & Associates, Inc.
San Antonio, Texas 78228

David Thissen, Ph.D. University of Kansas
Howard Wainer, Ph.D. Educational Testing Service
Michael Levine, Ph.D. University of Illinois
Malcolm Ree, Ph.D. USAF Human Resources Laboratory

The research here reported was carried out with the support of the
USAF Human Resources Laboratory, Contract #F41689-81-C-0012.

April, 1982

Standard Errors Associated with Item Response Theory Parameters

Abstract

The precision of methods associated with item response theory was investigated. Attention was concentrated on the three-parameter logistic model, but results were also found for two- and one-parameter models. The results presented here apply to maximum likelihood estimation and address the standard errors which are associated with the a , b , and c parameters.

Results show that in some commonly found cases, standard errors associated with parameters are so large that maximum likelihood estimation may not yield useful estimates unless sample sizes are inordinately large.

Introduction

The field of ability testing theory and practice known variously as Latent Trait Theory, Item Characteristic Curve Theory, and Item Response Theory (IRT) has developed with great rapidity during the last fifteen years. (Lord, 1980). The forward momentum of the field has led to new methods of equating, to progress in computer adaptive testing, and to increased measurement precision in diverse fields of testing.

Progress, however, has been uneven, and some theoretical areas have lagged behind the application areas. One such area is the development of estimates of standard errors and associated significance tests for parameters of IRT curves. The present research is directed toward that area.

The actual size of standard errors of items parameters will depend on a large number of factors. For example, if one is calibrating a test whose items range widely in difficulty, but one is using a sample of low ability examinees, then the item parameters for the difficult items will not be well estimated, because the large majority of examinees will give wrong answers to the difficult items.

Workers in the field of Item Response Theory know that it is difficult to compute maximum likelihood estimates of lower asymptotes of easy items, but the precise degree of uncertainty associated with that asymptote has not been well known until now.

The object of the present research has been to determine the size of standard errors for individual item parameters and for item characteristic curves as a whole, and to develop a method for testing whether two item characteristic curves are significantly different from each other. Owing to limitations of space, however, this presentation of results is limited to a discussion of the standard errors.

Method

By making certain simplifying assumptions, it is possible to develop complex, but accurate, functions to represent the standard errors of the item parameters. The assumptions included a normally distributed (0,1) sample of abilities, and a good fit between the model and data. It is also assumed that the abilities are known accurately. Since these assumptions will never be perfectly met, the tabled results represent lower limits (i.e. asymptotic values) of the standard errors obtainable in ideal circumstances. Expressions for the loglikelihood of the maximum likelihood estimates were found, as were the expressions for their second derivatives. The inverse of the negative expected values of the second derivative matrix is a matrix of the variances and covariances of (in this case) the item parameters, (i.e., standard errors). A computer program employing a multivariate form of Newton-Raphson estimation was written and used to evaluate the expressions and invert the matrices.

Results

Complete and detailed results are beyond the scope of this paper, but they are available in the technical report cited (Thissen and Wainer, 1981). The following tables are representative of the results for the three parameter model. In order to increase the utility of the tables, a general entry is provided, and to find a specific value for a given sample size, tabled values must be divided by the square root of the sample size. Table 1 gives standard errors for slopes (a), Table 2, for position (b), and Table 3, for lower asymptotes (c).

Analogous expressions were obtained for one and two parameter models.

Discussion

As an examination of the tables shows, the standard errors are in some cases large enough so that the estimations of the parameters contain less useful information than one would hope for. For example, with a sample of 10,000 examinees, the standard error of the position (i.e. difficulty, or b parameter) of an IRT curve (Table 2) at location -2 and slope 1.0 (representative of many operational items) is approximately 0.35. That means that the 95% confidence interval around the difficulty of such an item would stretch from -2.7 to -1.3. Clearly, users whose applications demand accurate estimations of such parameters must evaluate the consequences of the large uncertainty associated with those estimates.

References

- Lord, F.M. Applications of item response theory to practical testing problems. Lawrence Earlbaum Associates, Hillsdale, New Jersey, 1980.
- Thissen, D. and Wainer, H. Some standard errors in item response theory. Technical Report 81-12. Educational Testing Service, Princeton, New Jersey, 08541, 1981. (Parallel report to be released by USAF-HRL).

Table 1. Minimal asymptotic standards errors for slopes, three-parameter model, lower asymptote at 0.2.

Slope	Locations						
	-3	-2	-1	0	1	2	3
0.25	14.83	14.52	14.44	14.63	15.08	15.85	16.96
0.50	9.42	8.63	8.29	8.43	9.08	10.39	12.64
0.75	8.57	7.29	6.73	6.83	7.69	9.63	13.41
1.00	9.01	7.12	6.32	6.38	7.44	10.09	15.90
1.50	11.87	8.18	6.70	6.66	8.16	12.51	24.13
2.00	16.81	10.26	7.81	7.57	9.57	16.05	36.12
3.00	32.79	16.46	11.11	10.31	13.46	25.25	69.65

Table 2. Minimal asymptotic standards errors for locations, three-parameter model, lower asymptote at 0.2.

Slope	Locations						
	-3	-2	-1	0	1	2	3
0.25	878.74	705.10	569.00	459.07	366.54	284.60	207.77
0.50	210.96	136.60	92.05	63.17	42.23	25.60	22.18
0.75	111.23	59.47	34.93	21.68	13.36	10.10	21.56
1.00	76.42	34.95	18.63	10.97	6.90	7.82	20.38
1.50	48.07	17.87	8.51	4.88	3.60	5.98	18.14
2.00	35.36	11.76	5.31	3.09	2.62	4.92	16.32
3.00	23.72	7.10	3.06	1.88	1.86	3.71	13.31

Table 3. Minimal asymptotic standards errors for lower asymptotes, three-parameter model, lower asymptote at 0.2.

Slope	Locations						
	-3	-2	-1	0	1	2	3
0.25	94.54	72.66	56.80	45.21	36.65	30.28	25.49
0.50	50.27	29.36	18.15	11.92	8.32	6.17	4.84
0.75	42.68	19.62	10.06	5.76	3.67	2.59	1.99
1.00	40.61	15.39	6.87	3.60	2.18	1.51	1.16
1.50	38.53	11.27	4.27	2.04	1.19	0.83	0.66
2.00	35.92	9.13	3.20	1.47	0.87	0.62	0.51
3.00	30.56	6.92	2.28	1.05	0.64	0.48	0.43

Averaged Correlations Between Parallel Measures:
Reliability Estimation*

Alvah C. Bittner, Jr.
Naval Biodynamics Laboratory, New Orleans, LA.

Abstract

The utility of averaging sample and Fisher-z transformed correlations was investigated. Derived expressions revealed that averaging across (P) parallel measures markedly reduced the variance of estimated correlations. For reliability estimates, the variance reduction was greatest for (near) zero correlations and corresponded to reductions expected with increasing sample size by a factor of (nearly) $P(P-1)/2$. A limited study of cross-variable correlations, with a cross correlation of 0.50 and reliabilities of 0.71, indicated variance reductions by factors of 0.67 and 0.58 for selected examples with respectively two and three parallel measures. Overall, the findings of this report recommend the further investigation of averaged correlations and their practical application.

The sample product-moment correlation coefficient is a discouragingly variable statistic. Decreasing inversely with sample size (N), its variance is largest in the common case of weak correlation encountered in applied settings. Trattner and O'Leary (1980), in regard to sampling variation, have noted that it manifests itself in difficulties in establishing differences between correlations in differential validity studies. Similarly, Schmidt and Hunter (1976, 1977) have observed the deleterious impact of such variation on the problem of validity assessments. Historically, this variability has been taken to contraindicate multivariate statistical investigations (e.g., Factor Analysis) when sample sizes were small (Harmon, 1976). These authors altogether have concluded that, with the sample sizes ordinarily available, their considered applications of correlation coefficients are feasible less often than commonly assumed. However, utilizing multiple parallel measures could both reduce variability without increasing sample size and increase the feasibility of applications.

Use of multiple measures to improve assessment of relationships recently has been recommended by a number of authors (e.g., Sutcliffe, 1980; Epstein, 1980). The formulation of these recommendations lies on the classical Spearman-Brown formula, the relationship between the reliability of the sum of a number of parallel tests and the single test reliability (Ghiselli, 1964). By employing the sum of a number of measures of either a known criterion or predictor, the strength of a relationship can be increased and sample size limitations can be partially offset. The possible use of such an approach, it is noteworthy, was not touched upon by either Trattner and O'Leary (1980) or Schmidt and Hunter (1976, 1977). Alternatively to employing the sum of multiple measures, the stability of correlations would appear to be improved by "averaging" the

* This study was conducted under Bureau of Medicine and Surgery Work Unit No. MF58.525-00205027. The opinions are those of the author and do not necessarily reflect those of the Department of the Navy.

correlations between multiple parallel predictor and multiple criteria measures. Woodworth (1938) recognized this possibility for rank correlations and provided an erroneous equation for the standard error. More recently, Jones (1972) in his monograph on individual differences apparently recognized the averaging advantage; he averaged the sample intercorrelations of tasks which appeared to be differentially stable. However, neither of these authors explored the characteristics of averaging correlations.

The purpose of this report is to demonstrate the potential reduction in sampling variance obtained by "averaging" correlations based upon multiple parallel measures. Focusing upon test reliability estimation, the asymptotic variances of the simple and Fisher-z averages of a set differentially stable correlations measured on the same subjects will be derived. In light of these variances, the trade-off of sample size (N) and number of differentially stable measures (P) will be considered. Implications of the sample-size and repeated measures trade-offs for correlational studies will be discussed.

Method

Derivation of asymptotic variances will be done in two phases. In the first phase, a vector form for the $P(P-1)/2$ unique correlations in a P-variate matrix will be delineated. Three distinct values for element covariances will be derived from expressions given in Steiger (1980) and the frequency of each value in the variance-covariance matrix will be enumerated. Pertinently, the distinct values and their frequency are sufficient for determination of the variance of a mean correlation; this follows as the variance of the sum of a vector's elements equals the sum of its population variance-covariance matrix elements (cf., Morrisson, 1967, p 79). The second phase will combine the results into an expression for the asymptotic variances of the simple and Fisher-z averages. In the third and last stage, efficiencies for the simple and Fisher-z averages will be defined in terms of the variances of a single r or z value.

Correlation Covariances and Number

The sample correlation between P-variates on the same sample of (N) subjects contains $P(P-1)/2$ unique elements. These elements can be written as a vector which is asymptotically ($N \rightarrow \infty$) multivariate normal (Anderson,

$$\underline{r} = (r_{1,2}, r_{1,3}, \dots, r_{1,P}; r_{2,3}, r_{2,4}, \dots, r_{2,P}; \dots; r_{P-1,P})^t \quad (1)$$

1958). Denoting the asymptotic r_{ij} as ρ_{ij} and the asymptotic covariance between r_{jk} and r_{hm} as $\sigma_{jk,hm}$, Steiger (1980) has given the expressions

$$\psi_{jk}^2 = N\sigma_{jk}^2 = (1 - \rho_{jk}^2)^2 \quad (2)$$

$$\psi_{jk,jh} = N\sigma_{jk,jh} = \rho_{kh}(1 - \rho_{jk}^2 - \rho_{jh}^2) - \frac{1}{2}(\rho_{jk}\rho_{jh})(1 - \rho_{jk}^2 - \rho_{jh}^2 - \rho_{kh}^2) \quad (3)$$

$$\begin{aligned} \psi_{jk,hm} = N\sigma_{jk,hm} = \frac{1}{2}\{ & (\rho_{jh} - \rho_{jk}\rho_{kh})(\rho_{km} - \rho_{kh}\rho_{hm}) \\ & + (\rho_{jm} - \rho_{jh}\rho_{hm})(\rho_{kh} - \rho_{kj}\rho_{jh}) \\ & + (\rho_{jh} - \rho_{jm}\rho_{mh})(\rho_{km} - \rho_{kj}\rho_{jm}) \\ & + (\rho_{jm} - \rho_{jk}\rho_{km})(\rho_{kh} - \rho_{km}\rho_{mh}) \} \end{aligned} \quad (4)$$

where equation (2) applies if both subscripts are the same, equation (3) if one subscript is in common, and equation (4) if no subscripts are common. Under the assumption of differential-stability ($\rho_{ij} = \rho$ $i \neq j$), the equations (2), (3) and (4) simplify to only the three values shown in the second column of Table 1. These are the distinct elements of the asymptotic \underline{r} covariances matrix: $\Sigma_{\underline{r}}$.

TABLE 1: COVARIANCES BETWEEN SAMPLE AND TRANSFORMED CORRELATIONS VECTOR ELEMENTS AND THEIR NUMBERS

TYPE	CORRELATION COVARIANCES	TRANSFORMED FISHER-Z COVARIANCES	NUMBER *
BOTH SUBSCRIPTS OVERLAP	$N\sigma_1 = (1-\rho^2)^2 = \psi_1$	$(N-3)\sigma(z)_1 = N\sigma_1/(1-\rho^2)^2 = 1$	$P(P-1)/2 = Q_1$
ONE SUBSCRIPT OVERLAP	$N\sigma_2 = \rho(1-2\rho^2) + \rho^2(1-3\rho^2) = \psi_2$	$(N-3)\sigma(z)_2 = N\sigma_2/(1-\rho^2)^2$	$\left(\frac{P(P-1)}{2}\right)^2 - (Q_1 + Q_3) = Q_2$
NO SUBSCRIPTS OVERLAP	$N\sigma_3 = 2(\rho - \rho^2)^2 = \psi_3$	$(N-3)\sigma(z)_3 = N\sigma_3/(1-\rho^2)^2$	$\frac{P(P-1)(P-2)(P-3)}{4} = Q_3$

* NUMBER IN $(P(P-1)/2)^2$ Covariance Matrix

Table 1 also contains the distinct elements the asymptotic covariance matrix for the vector of Fisher-z transformed correlation: \underline{z} . Denoting the vector of individually z-transformed elements of \underline{r} by

$$\underline{z} = (z_{1,2}, z_{1,3}, \dots, z_{1,P}; \dots; z_{P-1,P})^t \quad (5)$$

where $z_{ij} = \frac{1}{2} \ln((1+r_{ij})/(1-r_{ij}))$, the third column gives the covariance for various levels of subscript overlap. These expressions were derived, substituting $\rho_{ij} = \rho$ $i \neq j$, from the relationships also given by Steiger (1980)

$$(N-3) \sigma_{z_{jk}}^2 = 1 \quad (6)$$

$$(N-3) \sigma_{z_{jk}, z_{jh}} = \psi_{jk, jh} / ((1-\rho_{jh}^2)(1-\rho_{jh}^2)) \quad (7)$$

$$(N-3) \sigma_{z_{jk}, z_{hm}} = \psi_{jk, hm} / ((1-\rho_{jk}^2)(1-\rho_{hm}^2)) \quad (8)$$

where $\sigma_{z_{jk}, z_{hm}}$ is the asymptotic covariance between z_{jk} and z_{hm} and expressions (6), (7), and (8) apply respectively for two, one and no subscript overlap. In Table 1, the analogous covariances for sample correlations and z-transformed correlations can be seen to differ by a factor proportional to $(1-\rho^2)^2$. This proportionality of the asymptotic covariances for sample and transformed correlations will simplify later derivations.

Table 1 enumerates the distinct covariance terms of the expected covariance matrices. Enumeration was made by noting that the $P(P-1)/2$ elements of the (\underline{r} and \underline{z}) vectors would have square covariance matrices with $(P(P-1)/2)$ rows and columns. Considering the number of unique covariances of elements of \underline{r} and \underline{z} with themselves, only $P(P-1)/2 = Q_1$ would be seen and these would be along the diagonal of the covariance matrix. The number of elements in each $\sqrt{\quad}$ with non-overlapping subscripts can be enumerated as $((P-2)(P-3)/2)$, yielding over the $(P(P-1)/2)$ rows a total of $Q_3 = (P(P-1)(P-2)(P-3)/4)$ matrix elements of this type. The number of row covariance elements with overlap of a single subscript can be derived as the differences of the total number of matrix elements and those with zero or two element overlap ($Q_2 = (P(P-1)/2)^2 - (Q_1 + Q_3)$). Table 1 summarizes both distinct covariance values and their numbers in the third column.

Variance of Averaged Correlations

The variances of the simple and z-transformed vector can be computed from values in Table 1. Indeed, the variance of the average (\bar{r}) of the $P(P-1)/2$ elements of vector \underline{r} is

$$\text{Var}[\bar{r}] = (P(P-1)/2)^{-2} [Q_1 \sigma_1 + Q_2 \sigma_2 + Q_3 \sigma_3] \quad (9)$$

where terms are defined as in Table 1. This result follows from the well-known result $\text{Var}(\underline{a} \underline{r}) = \underline{a} \underline{r} \underline{a}$ (e.g., Anderson, 1958; Morrison, 1967) with the covariance matrix for \underline{r} and \underline{a} defined as the $P(P-1)/2$ element vector with all elements equal to $(P(P-1)/2)$. A like expression can also be computed for the z-vector element average (\bar{z}).

$$\text{Var}[\bar{z}] = (P(P-1)/2)^{-2} [Q_1 \sigma(z)_1 + Q_2 \sigma(z)_2 + Q_3 \sigma(z)_3] \quad (10)$$

Significantly, it can be also written

$$\text{Var}[\bar{z}] = N(N-3)^{-1} \text{Var}[\bar{r}] / (1-\rho^2)^2 \quad (11)$$

because of the proportionality between covariance terms as seen comparing the second and third columns of Table 1. Expressions (9) and (11) will be those utilized in consideration of relative efficiencies.

Asymptotic Relative Efficiencies

Relative efficiencies of \underline{r} and \underline{z} can be defined in terms of the variances of single simple and z-transformed correlations based on the same sample size. For \underline{r} , the inverse efficiency ratio ($I_{\underline{r}}$) would be

$$I_{\underline{r}} = \text{Var}(\bar{r}) / [N^{-1} (1-\rho^2)^2] \quad (12)$$

where $\text{Var}(\bar{r})$ is defined by equation (9) and the denominator follows from (2). Likewise, the inverse ratio for \underline{z} would be

$$I_{\underline{z}} = \text{Var}(\bar{z}) / (N-3)^{-1} \quad (13)$$

where $\text{Var}(\bar{z})$ is defined by (10) and the denominator follows from (6). Substituting (11) into (13), it follows that

$$I_{\underline{z}} = \frac{N(N-3)^{-1} \text{Var}[\bar{r}] / (1-\rho^2)^2}{(N-3)^{-1}} = I_{\underline{r}} \quad (14)$$

defined in (12). This equivalence of $I_{\underline{r}}$ and $I_{\underline{z}}$ implies that only one set of inverse efficiencies ($I = I_{\underline{r}} = I_{\underline{z}}$) need be considered.

Sample Size and Repetition Trade-offs

The results of evaluating the efficiency ratio (I), showed that variances of averaged correlations were substantially reduced over the single correlation value in all cases. The maximum impact of increased repetition was observed with low correlations near or equal to zero. In general, for ρ at (or near) zero, the variance of the sample or Fisher-z average over P repetitions was (nearly) a factor of $[P(P-1)/2]^{-1}$ that for a single observation. Variation for $\rho = 0$ was, in particular, reduced by a factor of 1/6 for 4 repetitions, a reduction analogous to increasing the sample size by a factor of 6. This, and similar reductions, suggests the use of the average of multiple measures of the same subjects, vice large sample sizes as a method of enhancing correlation estimation precision.

Trade-offs of sample size and repetitions are restricted for the results presented in this study. The assumption of asymptotic sample size, underlying the derivations, may be appropriately applied only where N is "effectively asymptotic". Fortunately, the Fisher-z transformed sample correlations $z(r)$ can be expected effectively asymptotic results with even moderate N, a property not shared by the untransformed sample coefficient (Anderson, 1958; Rao, 1952). Examination of the moments for transformed sample bivariate normal correlations, given in Rao (1952), suggests that between $15 \leq N \leq 20$ will be adequate. Rao (1952), it is noteworthy, used this transformation over a set of correlations varying from 10 to 23 in his homogeneity test example. Hence the results given in this note would be applicable for Fisher-z transformed correlations based on sample sizes of the order of 20.

Extensions of the Present Results

In the present effort, the potential of averaging correlations was supported for reliability estimation. Applications to the problems outlined by Schmidt and Hunter (1976, 1977), and Trattner and O'Leary (1980), and other multivariate applications were not directly touched. However, the similarity of expressions to be expected in derivations following the present approach supports the potential for averaged cross variable correlations. Additional support for this potential comes from evaluating the variances for averaged z-transformed cross-correlations in limited cases. In particular, a brief investigation of averaging z-transformed cross variable correlations revealed results similar to those in Table 2. Assuming two variables with common and stable 0.71 reliabilities and 0.50 cross-correlation calculations based on equations (2), (3) and (4) were made. These calculations revealed that, with two parallel measures and averaging over the four transformed cross-correlations, the inverse efficiency is $I = 0.67$. Similarly averaging over the transformed cross-correlations, between three parallel measures of each variable, results in an efficiency of 0.58. Together, these results indicate that with respectively two or three parallel measures on each variable the variance of the correlation estimate is reduced analogous to increasing sample size by factors of 1.49 and 1.74 respectively. Certainly cross-variable validity studies appear to have enhanced feasibility when parallel measures and correlation averaging are employed.

Overall, the findings recommend the further investigation of averaged sample correlations and their practical applications.

Efficiency (I), it is noteworthy, is computationally independent of the sample size (N). This can be seen by, first, substituting ψN^{-1} terms defined in Table 1 for analogous σ terms into the $\text{Var}(\bar{r})$ expression (9) and, second, substituting the resulting expression into (12) to yield

$$I = (P(P-1)/2)^{-2} [Q_1 \psi_1^2 + Q_2 \psi_2^2 + Q_3 \psi_3^2] / (1-\rho^2)^2 \quad (15)$$

where all terms are only a function of P and ρ . This expression is the form which was employed in developing the results given in the following section.

Results

Table 2 contains the results of evaluating the inverse efficiency equation (15). Pertinently, the values of inverse efficiency (I) were computed with values of the repetition number (P) and correlation (ρ) as parameters. Examining this table, it can be seen that I is less than unity in all cases. Further, it can be seen that for any P, I increases as ρ increases from 0.00 to 0.99. The values of I, for any ρ , likewise decreases as the number of repetitions (P) increases from 3 to 50. The largest tabulated inverse efficiency ($I=0.747$) is seen when $P=3$ and $\rho = 0.99$ and the least tabulated value ($I=0.001$) is seen when $\rho = 0.00$ and $P=50$. Overall, these results demonstrate that the variances of \bar{r} and \bar{z} are substantially less than for their corresponding single values.

TABLE 2: INVERSE EFFICIENCY (I) AS A FUNCTION OF CORRELATION (ρ) AND NUMBER OF REPEATED MEASURES (P)

REPETITION NUMBER (P)	CORRELATION (ρ)											
	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	.95	.99
3	0.333	0.397	0.454	0.505	0.551	0.593	0.630	0.664	0.695	0.724	0.737	0.747
4	0.167	0.233	0.296	0.356	0.412	0.463	0.510	0.554	0.595	0.632	0.650	0.663
5	0.100	0.162	0.225	0.285	0.345	0.400	0.452	0.500	0.544	0.586	0.606	0.621
8	0.036	0.085	0.143	0.203	0.263	0.321	0.377	0.430	0.483	0.527	0.550	0.567
10	0.022	0.066	0.121	0.180	0.240	0.299	0.356	0.410	0.461	0.510	0.533	0.551
25	0.003	0.032	0.078	0.133	0.191	0.250	0.309	0.365	0.420	0.472	0.497	0.516
50	0.001	0.023	0.066	0.119	0.177	0.236	0.295	0.352	0.407	0.460	0.485	0.505

Discussion

This investigation was directed at derivation of the asymptotic variances of averaged correlations and consideration of trade-offs between sample size (N) and number of measures (P). Earlier, equations (9) and (11) respectively were found giving the desired variances for averaged sample (\bar{r}) and transformed (\bar{z}) correlations. Comparisons of these variances to that for comparative single values was provided for by derivation of the common inverse efficiency ratio I and calculation of values of I for various P and ρ given in Table 2. In the subsequent sections, trade-offs of P and N will be discussed and extensions of the present results will be considered.

REFERENCES

- Anderson, T. W. An introduction to multivariate statistical analysis. New York: Wiley, 1958.
- Epstein, S. The stability of behavior: II. Implications for psychological research. American Psychologist, 35, 1980, 790-806.
- Ghiselli, E. E. Theory of psychological measurement. New York: McGraw-Hill, 1964.
- Jones, M. B. Individual differences. In R. N. Singer (ed.), The psychomotor domain. Philadelphia: Lea and Febeger, 1972.
- Morrison, D. B. Multivariate statistical methods. New York: McGraw-Hill, 1967.
- Harmon, H. H. Modern factor analysis (3rd Edition). Chicago: University of Chicago Press, 1976.
- Rao, C. R. Advanced statistical methods in biometric research. New York: Wiley, 1952.
- Schmitt, F. L., & Hunter, J. E. Statistical power criterion-related validity studies. Journal of Applied Psychology, 1976, 61, 473-485.
- Schmitt, F. L., & Hunter, J. E. Development of a general solution to the problem of validity generalization. Journal of Applied Psychology, 1977, 62, 529-540.
- Steiger, J. H. Tests for comparing elements of a correlation matrix. Psychological Bulletin, 1980, 87, 295-251.
- Sutcliffe, J. P. On the relationship of reliability to statistical power. Psychological Bulletin, 1980, 88, 509-515.
- Trattner, M. H., & O'Leary, B. S. Sample size for specified statistical power in testing for differential validity. Journal of Applied Psychology, 1980, 65, 127-134.
- Woodworth, R. S. Experimental psychology, New York: Holt, 1938.

Development of the Common Metric

Bennie W. Roach and Deborah L. Rogers

Air Force Human Resources Laboratory
Brooks Air Force Base, Texas 78235

Abstract

The Air Force Officer Qualifying Test (AFOQT) has been used for selecting officers into precommissioning and technical training programs for 30 years. The normative base used for standardizing the test battery has changed several times during its operational lifetime. An evaluation indicated that although recent successive forms of the AFOQT were anchored to the same normative base, they were not equated to each other. The purpose of this study was to establish a new normative base and to establish a method for placing the results of all applicants taking recent, current, and future forms of the AFOQT on the same scale.

The first major instrument for selecting Air Force officers was the Aircrew Classification Battery (ACB) developed during World War II by the Aviation Psychology Program. But, the end of the war signaled an end to the huge demand for aircrew members, so the testing program was curtailed (Flanagan, 1948). A revived interest in selection test development started in the fall of 1949 with a requirement for an Aviation-Cadet Officer-Candidate Test (AC-OC-AT).

The first explicit use of the name "Air Force Officer Qualifying Test (AFOQT)" in designating a set of test booklets occurred with a preliminary version constructed in 1951. This version incorporated the AC-OC-AT and was designed to yield scores predictive of success in the Officer Candidate School and in officer technical courses. The first operational AFOQT evolved from this preliminary battery (Valentine & Creager, 1961).

In the early fifties the AFROTC program was reoriented to become a major source of flying personnel for the Air Force. Therefore, most selection testing was performed at the AFROTC detachments. The first form of the AFOQT (Form A) was administered to AFROTC cadets in 1953 and 1954, and was a simplified version of the Aircrew Classification Battery. Unlike the ACB, it did not contain the well-known psychomotor apparatus tests. Although psychomotor tests would have added validity to the paper-and-pencil portion of the battery, they would have been too expensive to administer at the 200 AFROTC detachments, which were in operation in 1953. A second difference was the omission of any speeded subtests. This, too, was expected to result in the loss of some validity, but this outcome had to be weighed against the possibility of serious timing errors in the administration of highly speeded tests by relatively untrained examiners.

Since 1954, a new form of the AFOQT has been produced every 2 to 3 years. The later editions were similar to Form A except that speeded subtests were reintroduced. The present operational version, Form O, is number 15 in the AFOQT series.

During the past 4 decades of officer selection and classification the reference group used for norming purposes has changed many times. The Aircrew Classification Battery, as well as most of the test batteries developed by the Aviation Psychology Program, used the World War II mobilization population as a norming base. The end of the war resulted in a loss of this population as a reference group.

When officer selection testing resumed in 1949 the norming base was composed of West Point cadets. But, with the opening of the Air Force Academy (AFA), a decision was made to change the norming base again. By 1955, the AFOQT conversion tables were based on the AFA applicant population (Valentine & Creager, 1961). The norming procedure was soon modified, however. An evaluation of performance by the applicant group on the College Entrance Examination Board (CEEB) indicated that successive applicant groups tended to become more self-selective with respect to aptitude. In order to maintain comparability of scores between forms, the CEEB was used as an instrument to link successive forms of the AFOQT. This was accomplished by establishing an equipercentile conversion between the CEEB and the AFOQT for each class normed.

The inclusion of both the AFOQT and the CEEB in the Academy selection process created a problem of redundancy in testing. Since the CEEB provided the advantage of permitting direct comparison of Academy groups with other college level groups, the requirement for the AFOQT as a selection test was eliminated by the Air Force Academy in 1960 (Miller & Valentine, 1964).

In anticipation of the loss of the AFA applicant population as a reference group, a new normative base was obtained by administering Form G of the AFOQT and the Project TALENT test battery simultaneously to more than 5,000 applicants for the AFA class of 1964. Both test batteries were also administered to about 1,000 basic airmen. By utilizing the relationships between the AFOQT and the TALENT composites, an indirect process of standardizing each new form of the AFOQT was established. This procedure is described in detail by Dailey, Shaycoft, and Orr (1962) and Miller (1969).

The indirect method of using TALENT composites and basic airmen became the accepted procedure for standardizing successive forms of the AFOQT. Although this standardization procedure provided a means of relating successive forms of the AFOQT to an AFA applicant group, there were no attempts made to make successive forms comparable with each other. The belief that the indirect method was completely adequate is well documented in the literature. For example, according to Gregg (1968), "Successive forms of the AFOQT are identical in terms of psychological functions measured. Differences between successive forms are largely a matter of item content and form and minor administrative adjustments. Thus, it is assumed that an individual will achieve similar scores if he takes two forms of the test at about the same time" (p. 1). And Miller (1969) stated that, "Since each AFOQT form is standardized by referring back to the original TALENT composite distributions, an unchanging normative base is achieved which permits direct comparisons of scores on successive AFOQT forms. The stratification of the standardization groups permits comparison of any AFOQT composite with any other" (p. 23).

Gregg (1968) also pointed out that, "Three to four years of collegiate experience coupled with the concomitant maturation have a marked elevating effect on scores obtained on the AFOQT" (p. 11). As a result of this study, two conversion tables were introduced to account for different educational levels. The conversion table for examinees with less than 2 years of college was designed for AFROTC applicants, who were normally administered the test during their freshman or sophomore year. The college graduate conversion table was designed for OTS applicants who tested near graduation from college. In 1972 the number of conversion tables was increased to three in order to provide a more accurate evaluation of those examinees who would take the AFOQT during their junior or senior year of college.

During the mid-1970's there was a marked increase in the number of females applying for precommissioning programs. Since the officer and pilot biographic inventories had not been normed on females, they were not administered to female examinees. Thus, additional conversion tables were added in 1974. By the time Form M was implemented, multiple conversion tables were in effect for each form of the AFOQT. The appropriate conversion table was determined by the sex and educational level of each examinee.

In an attempt to improve the validity of the AFOQT, Valentine (1977) performed an extensive validation study. Form N, which was implemented in 1978, was significantly revised as a result of this study. Since many of the subtests contained in previous versions of the test battery had been replaced, it was impossible to provide equipercentile conversions using the previously established indirect method. The new normative base was designed to be representative of all major sources for Air Force commissioning programs and sources for specialized training programs. The reference group of almost 2,700 cases was composed of basic airmen; AFROTC, OTS, and AFA cadets; and junior officers. The sample was then divided into three subgroups based on the education level of each examinee at the time of testing. Separate conversion tables predicted on sex were not required with Form N since the Officer Biographic Inventory had been dropped and the Pilot Biographic Inventory was normed on both males and females. Gould (1978) provides a detailed description of the standardization of Form N.

The Form N norming sample differed from previous groups used to norm the AFOQT in two important aspects. First, the sample was not linked to the AFA class of '64. Second, the sample was a blend of potential applicants, selectees for precommissioning programs, and actual commissionees. Previous normative bases had been limited to cases representative of the potential applicant group only.

By 1980 the AFOQT scores used for selection into officer training programs were no longer comparable between groups. Since there was no time limit on how long AFOQT scores would be considered valid, an individual could apply for a precommissioning program with scores that were up to 10 years old. Therefore, an applicant's percentile score could be determined from any one of the 18 conversion tables obtained from four different versions of the test. A score depended on the sex of the applicant, and the education level and AFOQT form in use at the time of testing. For example, suppose three applicants apply for OTS at the same time and all three performed equally well on the Officer Quality composite of the test battery, but the sex, education level, and form of the test differed. Applicant A is a male who took Form L of the AFOQT in 1972 as a college freshman. Applicant B is a female who took Form M in 1976 as a college senior. And, applicant C is a male who took Form N in 1979 as a college graduate. The percentile scores obtained were 40 for A, 25 for B, and 50 for C. This means that applicant A performed better than 40% of all male college freshmen taking Form L, applicant B performed better than 25% of all female college seniors taking Form M, and applicant C performed better than 50% of all college graduates taking Form N of the AFOQT. While it was possible to determine the relative standing of each applicant within their respective peer group, it was not possible to accurately rank-order these applicants between groups. While all three applicants were of equal ability, applicant C would be ranked higher than applicant A, who was ranked higher than applicant B.

In other words, as long as the selection process occurred in an idealistic setting with a homogeneous group of applicants, the scores would be meaningful. But in a realistic selection environment, homogeneous applicant groups do not normally exist. Therefore, the purpose of this study was to establish a method for placing the results of all applicants taking recent, current, and future forms of the AFOQT on the same scale. This scale, referred to as the common metric, allows for an accurate comparison of individual and group scores.

Method

Standardization of the AFOQT must provide a uniformity in the meaning of test scores without regard to the time or place of testing. This uniformity requires that a single reference group be the basis of all conversion tables. Therefore, the development of a common metric is a two-phase process. Phase I is to develop a suitable reference group which can be adequately described and will serve as a stable anchor for an extended period of time. Phase II is the development of the single scale which will be used to provide the actual standardization of test scores.

An appropriate reference group should satisfy five basic prerequisites. First, the group must be adequately defined in order to be a standard for reference; e.g., the World War II mobilization population, the applicants for the Air Force Academy class of 1964, etc. Second, the group must be relevant to the intended purposes of the selection process. Third, sampling error must be minimized. The larger the sample the more likely this requirement will be satisfied. Fourth, the group must be representative of the intended testing population. If the sample is biased due to disproportionate representation of subjects on demographic variables such as sex, race, or source of subject, potentially serious error may be introduced. Finally, the influence of motivation must be controlled. Often the environment of an experimental testing situation used to collect data does not encourage the individual to perform at his or her full potential. For example, an examinee that has already been accepted into a precommissioning program may be less motivated on a selection battery administered for "research purposes."

In 1980 an effort began at the Air Force Human Resources Laboratory (AFHRL) to implement a centralized scoring system which provided electronic scanning of the answer sheets and computerized scoring of item responses. Since AFOQT test records had been sent to AFHRL for archival storage, it was possible to build a very large data base containing test records on individuals given the AFOQT at non-ROTC test sites since January 1975. AFROTC had implemented a similar centralized scoring system several years earlier. A data base containing the results of AFOQT testing at the ROTC detachments since January 1972 was also available at AFHRL. By combining the two sources of data it was possible to build a consolidated data base which included the results on about 145,000 individuals who took a recent version of the test battery in an operational environment.

The consolidated AFOQT data base appeared to be an ideal source for a new normative base. The group was certainly relevant to the intended purposes of the selection process. Since it included essentially the entire testing population, the sampling error would be eliminated and the group would be unquestionably representative. And, the influence of motivation was adequately controlled since the test records were obtained from operational testing environments. In an operational environment, examinees are taking the test for the purpose of applying for a precommissioning or

specialized training program. Therefore, it can be generally assumed that the examinee tested in an operational environment will be more highly motivated than one tested in a research environment, and the scores will be a more realistic measure of true ability.

The only significant limitation to use of the total group was the fact that it contained examinees who had taken three successive, but not equated, forms of the AFOQT. Therefore, individuals who had taken only one form of the test could be used.

Form L of the AFOQT was administered from 1972 to 1975, but test records were not available on individuals who tested on Form L at non-ROTC test sites prior to January 1975. Test records were available on almost everyone who took Form M from 1975 to 1978. But this period in time was atypical for Air Force officer selection for two reasons. First, the end of the Vietnam era had produced a surplus of officers on active duty. As a result, the OTS precommissioning program was all but closed. Second, the end of the draft and the pervasive anti-military sentiment of the mid-70's was suspected of having had an adverse influence on the characteristics of the AFROTC applicant population.

Test records were also available on essentially everyone who took Form N since April 1978. The surplus of officers from the Vietnam era had been greatly reduced and OTS had increased its input. The anti-military influence was fading from the campuses, and the nation was in a period of relative economic stability. Therefore, the group of applicants taking Form N of the AFOQT at operational test sites between April 1978 and September 1980 is defined as the reference group for the new normative base.

Once the normative base was established, Phase II of the standardization process was simple and straightforward. The distribution of test performance was determined on each of the five composites (Pilot, Navigator-Technical, Academic Aptitude, Verbal, and Quantitative) of the AFOQT for the almost 60,000 applicants making up the reference group. The distribution for each composite became the common metric for that respective composite. Next, it was necessary to link Form M to the reference group. Since a certain portion of the total test items from one form of the test are used on a succeeding form, the common items between forms provided a means for linking adjacent forms. Through equipercentile converting, Form M was linked to Form N and placed on the common metric. In a similar fashion Form L was linked to Form M and then placed on the same common metric. Form O, which became operational in August 1981, was linked to the reference group and placed on the common metric as a part of test construction.

Results and Comments

As a result of this study, two significant improvements to the AFOQT were implemented. First, a new normative base was established. This reference group replaces the 20-year-old group of applicants for the AFA class of '64, and is referred to as the "Post-Vietnam Anchor Group." And second, a single scale, referred to as the common metric, was developed which provides uniformity in the meaning of AFOQT test scores.

The major advantages of using the common metric are:

1. Four adjacent forms (L, M, N, and O) of the AFOQT are directly linked together and placed on the same scale.
2. Examinees tested over almost a decade of officer selection can be compared, either as individuals or as groups.
3. Test performance is reported as a percentile score referenced against a population of actual applicants for officer training programs.

4. Air Force decision makers have a more accurate means of evaluating trends in applicant flow, the efficacy of selection programs, and the results of changes in policy.

REFERENCES

- Dailey, J. T., Shaycoft, M. F., & Orr, D. B. Calibration of Air Force selection tests to Project TALENT norms. PRL-TDR-62-6, AD-285 185. Lackland AFB, TX: Personnel Research Laboratory, Aerospace Medical Division, May 1962.
- Flanagan, J. C. The aviation psychology program in the Army Air Forces. Research Report No. 1. Washington, D. C.: U.S. Government Printing Office, 1948.
- Gould, R. B. Air Force Officer Qualifying Test Form N: Development and standardization. AFHRL-TR-78-43. AD-A059 746. Brooks AFB, TX: Air Force Human Resources Laboratory, Personnel Research Division, August 1978.
- Gregg, G. The effect of maturation and education experience on Air Force Officer Qualifying Test scores. AFHRL-TR-68-107. Lackland AFB, TX: Air Force Human Resources Laboratory, Personnel Research Division, July 1968.
- Miller, R.E. Interpretation and utilization of scores on the Air Force Officer Qualifying Test. AFHRL-TR-69-103. AD-687 089. Lackland AFB, TX: Air Force Human Resources Laboratory, Personnel Research Division, May 1969.
- Miller, R. E., & Valentine, L. D., Jr. Development and standardization of the Air Force Officer Qualifying Test-64. PRL-TDR-64-6, AD-600 782. Lackland AFB, TX: Personnel Research Laboratory, Aerospace Medical Division, March 1964.
- Valentine, L. D., Jr. Navigator-observer selection research: Development of new Air Force Officer Qualifying Test Navigator-Technical composite. AFHRL-TR-77-36. Brooks AFB, TX: Air Force Human Resources Laboratory, Personnel Research Division, May 1977.
- Valentine, L. D., Jr., & Creager, J. A. Officer selection and classification tests: Their development and use. ASD-TN-61-145, AD-269 827. Lackland AFB, TX: Personnel Laboratory, Aeronautical Systems Division, October, 1961.

The Use of Conjoint Analysis
in the Interval Subjective Scaling of Mental Workload

Stephen P. Boyd

Virginia Polytechnic Institute and State University

Abstract

Conjoint analysis is a technique which has been applied to subjective scaling of mental workload, with the goal of deriving an interval scale. Three proposed dimensions of mental workload have been used to construct a matrix; subjects then rank the matrix cells according to overall mental workload. The method used to derive the overall rank order suppresses information regarding intersubject variation. This study produces evidence that subjects were unable to reliably order the matrix cells. Thurstonian scaling was used to cross-check the conjoint scale. Results indicate that there is insufficient derivation of an interval scale using conjoint analysis.

The Use of Conjoint Analysis
in the Interval Subjective Scaling of Mental Workload

A great deal of study has been devoted to the search for a subjective metric of mental workload (w/l) which yields an interval scale. Many researchers agree that w/l is a complex, multidimensional construct. Sheridan and Simpson (1979) proposed three possible dimensions of mental w/l. Time loading corresponds to the contribution of time constraints and interruptions to overall mental w/l. Effort loading deals with the complexity/uncertainty of the tasks. Stress loading comes from confusion, frustration, or anxiety associated with the task.

If one accepts the validity of the three dimensions of mental w/l described above, then conjoint analysis provides a means of deriving a workload scale having interval characteristics. Krantz, Luce, Suppes, and Tversky (1971) thoroughly describe the mathematics, which require ordinal discriminations. General descriptions of three levels for each dimension would be provided (Reid, Shingledecker, and Eggemeier, 1981). A 3X3X3 matrix represents all possible combinations. The subject then rank orders the cells of the matrix, according to the expected level of overall mental w/l.

The first step in the analysis is to determine the mathematical model which best describes the manner in which the subject combines the dimensions to arrive at the rankings. These axiom tests are described by Krantz and Tversky (1971). After the model is defined, scale values are manipulated to achieve a best fit with the rankings.

It may be possible to take advantage of intersubject variations by using Thurstonian scaling, which utilizes ranking variations in the derivation of scale values (Thurstone, 1927). According to Thurstone's Law of Comparative Judgement, the pairs of matrix cells whose ranks are frequently transposed are closer along the psychological continuum than those whose rank order are seldom or never transposed. If there is intersubject variation in the rankings, a Thurstonian analysis will indicate the relative psychological distances between the matrix cells. Such an analysis will provide a statistical cross-check of the conjoint analysis.

Method

Subjects

Thirty four male and female subjects performed the ranking task. All were undergraduate students at Virginia Polytechnic Institute and State University, between the ages of 18 and 25 years, and participated voluntarily in the study.

Procedure

First, a description of each dimension was provided, with examples illustrating the general effects of the

dimensions on mental w/l. After the explanations, the subjects were given instructions concerning how to accomplish the ranking of the w/l matrix cells. The rankings were accomplished using index cards. Each cell had a corresponding card, on which was printed descriptions of the appropriate levels of each dimension. The cards were given to the subjects in random order. Subjects simply arranged the cards in a stack, according to their prediction of overall w/l, in a three step procedure developed by Reid, et al (1981). The rankings were averaged across all subjects, mean ranks and resulting integer ranks were computed and entered into the conjoint analysis. Thurstonian analysis was also conducted on the ranks.

Results

Figure 1a shows the results of the ranking of the w/l matrix cells by the participants. Each cell can be represented by a three digit number, corresponding to the levels of the three dimensions. In each case, a "1," "2," or "3" represent the low, medium, or high level of a dimension, respectively. The dimensions are, in order, time loading (T), effort loading (E), and psychological stress (S) loading.

insert Figure 1 about here

The first item of interest is the variation in the ranks. The standard deviations of the ranks ranged from zero (0) for cell 111 to 5.00 for cell 321. In the latter case, plus or minus two standard deviations would include an interval of 20 ranks, or 74% of the entire scale. Only four of the 27 cells have a s.d. of less than 2.0. Such an extreme amount of inter-subject variation in the ranks should make one question the assumption that central tendency (mean) is a reliable method for establishing the rank order of the w/l cells.

The average ranks fall into seven distinct groups. The significance of the grouping becomes apparent if one sums the digits representing the levels of the three dimensions ($T+E+S=3,4,5,6,7,8,9$). The ΣTES value is constant within each group. For example, the cells 311, 212, 221, 131, and 122 ($\Sigma TES=5$) have mean ranks which are very closely spaced.

The fact that ΣTES is constant within groups suggests that the weights associated with each level of the three dimensions are all approximately equal. There appears to be little difference in average ranks, as long as the TES is held constant. This implies that a simplistic additive process underlies the rankings as performed by the subjects. The strength of the groupings is reinforced if one computes the average distance between adjacent cells. The average

distance between cells within groups is 0.41, while the spacing between groups averages 2.95. Thus, cells within groups tend to be 7.26 times as closely spaced as those adjacent cells in different groups. Kendall's coefficient of concordance (W) was computed for the rankings ($W = 0.83$).

A computer program was written to calculate Thurstone scale values based on the method of rank order. The resultant scale values for all cells are shown in Figure 1b.

The most important feature of the Thurstone scale values is that the groupings are preserved. It should be remembered that the average ranks are an estimate of distribution location, while the Thurstone scale is based on dispersion. This means that the cells within groups were much more frequently transposed in rank order than those between groups. The average distance between cells within groups is 2.95, while the between-groups average is 23.5, yielding a ratio of 7.96. This ratio is very close to the one derived from average ranks. Thus, the Thurstone scale values are comparable with the average ranks, with respect to both the presence of the groups, and in their relative spacing.

Next, a conjoint scale representing the cells of the $w/1$ matrix was calculated. The CONJOINT algorithm executed axiom tests, and determined that an additive model was appropriate for the data. The NOMETR algorithm then determined scale values using an additive model, yielding a stress level of $\theta = 0.0134$. These scale values are shown in Figure 1c. The groupings present in the average ranks and the Thurstone scale are essentially nonexistent in the conjoint scale.

Discussion

To evaluate the appropriateness of the conjoint scaling technique from a statistical point of view, there were three sets of data to compare. The summary statistics for the raw data give the first indications that there may be problems with the method employed to derive the conjoint scale. Foremost among the indications is the extreme amount of intersubject variation in the rankings. If one excludes cells 111 and 333, the average s.d. is 3.24, while the average spacing within groups was 0.406. Thus, the order of the cells within groups remains suspect.

As stated before, the Thurstone scale values are based upon how frequently one cell is judged greater than another. The more consistent the order, the greater the subjective distance between the cells. The scale values are derived from overlap of the ranking distributions, rather than from central tendency. However, the Thurstone scale nearly duplicates the grouping of the average ranks, providing strong corroborating evidence. Therefore, one cannot reasonably conclude that there is a well defined ordering within groups. The ranking process has yielded only seven

consistent ordinal discriminations, instead of 27. The conjoint scale is based completely upon the order in which all 27 cells are ranked. The use of mean ranks to derive the overall rank order suppresses all dispersion information, and may in fact merely introduce an artificial rank order within the groups, severely damaging the validity of the interval scale produced by conjoint analysis. Thus, the conjoint algorithm ignores the significance of intersubject variation. The data suggest that the subjects were reliably discriminating only seven ordinal levels of mental w/l.

Reid et al. (1981) used Kendall's coefficient of concordance to justify their use of average ranks. However, the W statistic compares all possible pairs of cells, and thus may be overpowered by the strong general ordinal trend of the groups. In other words, this statistic is too coarse to adequately sense transposition of neighboring cells. When there is a great deal of variation in the ranks, one is justified in questioning even the ordinal strength of the data. In this study, a large amount of such variation was, in fact, measured. It therefore seems that the subsequent derivation of interval data is extremely tenuous, if one uses central tendency alone as the method for modeling the ranking process.

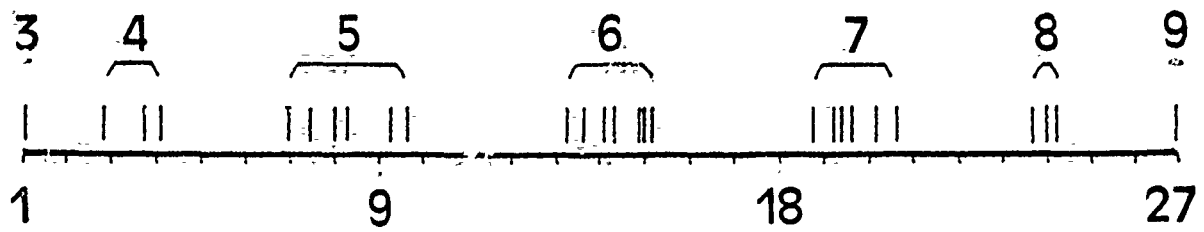
One can only speculate as to why the w/l discrimination task produced an ambiguous rank order. Possibly, variations occurred because the stimuli were ranked only indirectly. The index cards represent generic situations; the subjects must rank them according to what they think they would experience under the conditions described on the cards. The psychological disparity between the verbal descriptors and the actual experiences may contribute to the confusion. Another possibility is that people experience the combinations of the dimensions in very different ways, and that the groups are merely a result of summing across subjects. On the other hand, the rankings may reflect reality: that only seven levels of these dimensions are discriminable, even in actual experience. In any case, a seven-level rather than a 27-level ordering has been produced. This does not imply that conjoint analysis has no utility in the scaling of mental w/l. Rather, it suggests that either a smaller matrix and/or different dimensions and descriptors should be explored.

References

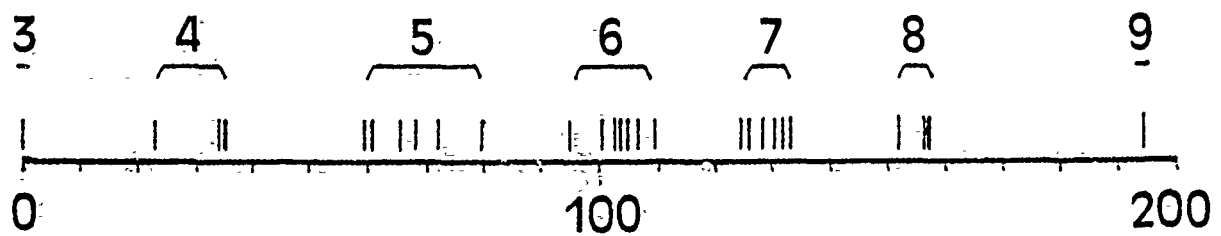
- Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. Foundations of measurement. New York: Academic Press, 1971.
- Reid, G. B., Shingledecker, C. A., & Eggemeier, F. T. Applications of conjoint measurement to workload scale development. Paper presented at the meeting of the Human Factors Society, Rochester, October 1981.
- Sheridan, T. B., and Simpson, R. W. Toward the definition and measurement of the mental workload of transport pilots (Contract No. DOT-OS-70055). Cambridge: Massachusetts Institute of Technology, January 1979.
- Thurstone, L. L. A law of comparative judgement. Psychological Review, 1927, 34, 273-286.

Figure Captions

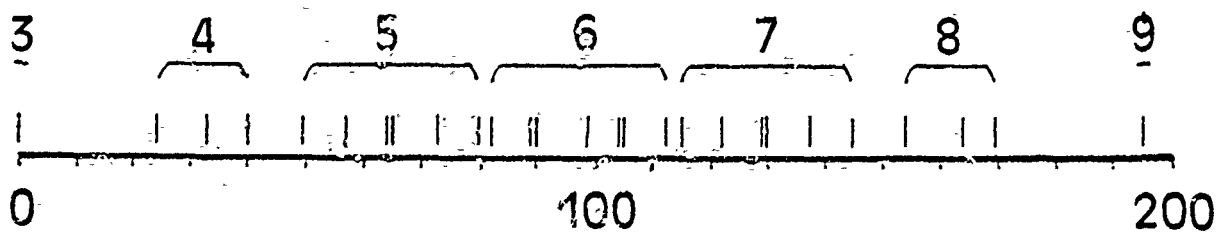
- 6a. Mean ranks of workload matrix cells
- 6b. Thurstone scale values
- 6c. Conjoint scale values



average rank



Thurstone scale value



conjoint scale value

PSYCHOLOGISTS IN THE AIR FORCE:
UTILIZATION, STRUCTURE, AND JOB ATTITUDES

Lt Col J. L. Mitchell & Mr. James B. Keeth
Occupational Analysis Program
USAF OCCUPATIONAL MEASUREMENT CENTER
Randolph Air Force Base, Texas 78150

and

Capt Linda A. Wiekhorst
Industrial Psychology Program
St. Mary's University
San Antonio, Texas 78228

ABSTRACT

This paper will discuss the results of an occupational analysis of Air Force Specialty 267X, Behavioral Scientist, positions in the United States Air Force (USAF). This study surveyed 80 percent of all USAF Behavioral Scientists and a few members of related specialties in terms of the tasks they perform, their qualifications, their attitudes about their jobs, and their career intentions. Job data were hierarchically clustered on the basis of similarity of tasks performed and the relative time spent on tasks. Resulting job groups are discussed in terms of personnel utilization, the implications for career structure, and attitudes of respondents. Several potential problem areas are identified.

INTRODUCTION

The United States Air Force (USAF) has a long history of successful behavioral science research. A number of lines of research developed during and following World War II included pilot selection, equipment design (Human Factors Engineering), animal research in space (Experimental Psychology), and various personnel and training programs (Personnel Tests & Measurements). In 1954, the USAF classification system recognized four types of behavioral scientists: AFS 8696, R&D Officer; 8854, Human Resources Staff Assistant; 8836, Human Resources Research Officer; and 8816, Human Resources Staff Officer. By 1961, these had become: 1896F, Human Performance Engineer; 2969E, Experimental Psychologist; 2696F, Personnel Measurement Psychologist, and 2616, Behavioral Scientist. In 1964, the initial three were converted into one field, 2675, Behavioral Scientist with shredouts: A for Human Factors, B for Experimental, C for Personnel Measurement, and a fourth shredout, Z, for related areas such as Physical Anthropology, Sociology, etc. In 1976, the shredouts were dropped to provide more flexible assignments. Since that time, there have been a number of suggestions to return to a shredout structure or to transfer some groups, particularly the Human Factors Psychologists, to some other Air Force specialty.

Throughout this period, there have been some suggestions that military psychologists were difficult to attract and retain. A 1968 report by the ad hoc committee on career status of military psychologists (Division 19, American Psychological Association) concluded that pay and allowances of uniformed psychologists compared "very unfavorably" with their civilian peers (Hedlund et al 1968). Salary was not the only concern, however, since job satisfaction was also thought to be impacted by the need for "opportunity for professional or scientific development," "more personal control over assignments," and by "professional or scientific isolation" (Ibid: 121). This report was followed in 1970 by an article by Dr Jacob Jacoby (1970) critiquing military salaries and utilization policies. Jacoby recommended entry at the Captain level for psychologists with doctorates, professional pay comparable to physician and flight pay, preinduction counseling on job opportunities, and a program of professional interaction through periodic conferences (Ibid: 386). Jacoby also challenged the Military Psychology Division (Division 19) of the American Psychological Association to take a more active role in lobbying for proper treatment of psychologists and dissemination of information about military psychology jobs (Ibid: 387).

Little research on these issues has been done or reported during the subsequent decade. While Division 19 has been developing a brochure of information on military psychology jobs, their publication requires further refinement before distribution (Wiskoff, personal communication 1981).

OCCUPATIONAL SURVEY OF USAF BEHAVIORAL SCIENTISTS

An occupational survey of USAF Behavioral Scientists was proposed by the USAF Occupational Measurement Center (USAFOMC) staff in 1978, as a part-time project which could be accomplished with minimal expense along with other projects. The relatively small utilization field population would not justify a normal priority project. The USAF Classification Branch of the Air Force Manpower and Personnel Center (AFMPC) concurred with this approach and approved of the project.

A USAF Job Inventory for Behavioral Scientists was developed using a set of special job descriptions collected by the Air Force Human Resources Laboratory (AFHRL) during 1974-1975 for another purpose, through interviews with one-third of current 267X job incumbents, and by telephone contact with a number of senior behavioral scientists. A relatively general, short task list was developed containing 330 tasks grouped under 17 major duty headings. Duties included General Command Functions (A), Supervisory Functions (B), Administrative Functions (C), General Functions (D), and a separate duty for each of the more specialized functional areas (for example, Contract Monitoring, Human Factors Engineering, Instructing, Test Development, etc). Also included in the USAF Job Inventory was a fairly extensive background section covering educational requirements, experience, job interest, special experience identifiers, prior AFSCs, and other information. The final job inventory was validated through comprehensive reviews by senior behavioral scientists at AFMPC, AFHRL, and USAFOMC. In addition, the AFMPC Career Development Manager also reviewed the instrument.

FIELD ADMINISTRATION

The USAF Job Inventory was mailed directly to organizations utilizing AFS 267X officers, and to those individuals in one-deep positions or at remote locations. Current membership of the utilization field was determined by a roster furnished by the AFMPC Career Development Manager. The inventory was mailed to the field in May and June 1981, with a request for return by mid-July. In late July, follow-up telephone calls were made to units and individuals who had not yet responded. Survey administration was closed in August 1981, and data entry begun. Mr. Bill Feltner served as CODAP programmer for the study and by 1 October had produced the initial CODAP analysis product. In addition, a newly developed program to identify core tasks for each job group (CORESET) was used (see Phalen and Weismuller 1981).

SAMPLE

Inventory booklets were received from a substantial cross section of the agencies and organizations which utilize behavioral scientists (see Table 1). The Air Training Command has the most behavioral scientists assigned and is most represented in the sample. Air Force Systems Command is the second largest user of behavioral scientists, including some one- or two-deep Human Factors and Research Manager positions. To ensure that a complete picture was achieved, a small number of personnel in related specialties (Education and Training Staff, Instructor, Commander, etc.) who hold a 267X secondary AFSC were also included in the study.

Generally, the sample is very representative of the utilization field, with about 80 percent of all behavioral scientists (or 113 of the 141 assigned) included (see Table 2). A few jobs or locations are not included. Several individuals retiring or leaving the USAF declined to respond, and in one case all three incumbents in an organization went PCS early in the summer, with no overlap with their replacements. While the lack of these members could lead to some sample bias, the amount of distortion was considered to be minimal since 80 percent of the field is captured in the sample.

RESULTS

One of the major objectives of the study was to identify major types of jobs, representing various ways behavioral scientists are utilized in the Air Force. Using the new CORESET program developed by Mr. Bill Phalen and Mr. Johnny Weissmuller, an analysis was made of the initial 38 groups appearing on the hierarchical clustering diagram. Using this new technique, we were able to collapse these 38 starter groups into 25 job types which grouped into seven major clusters. These included: Research Programs Scientists (35% of the sample); Functional Unit Supervisor (10%); Academic Instructors-Counselors (10%); Junior Task Scientists and Students (9%); Occupational Analysts (15%); Human Factors Engineers and Researchers (7%); and Test Development Psychologists (12%). These major clusters of jobs include 98 percent of the sample; the remaining individuals were filling one-deep, unique positions.

Table 1
ORGANIZATIONS REPRESENTED

	NO. IN SAMPLE
<u>AIR TRAINING COMMAND</u>	
HQ DCS PLANS (6), TECHNICAL TRAINING HQ & CENTERS (7), AIR UNIV (LMDC/AWC-4), AFIT STAFF & STUDENTS (5), AFROTC (3), USAF RECRUITING SERVICE (1), USAFOMC (41)	67
<u>AIR FORCE SYSTEMS COMMAND</u>	
AFHRL (12), HQ (3), USAFSAM (1), ESD (1), FTC (3), ASD (3), SAMSO (2), AFAMRL (2), AFWL (1), OTHER (3)	31
<u>USAF ACADEMY</u>	22
<u>DOD AGENCIES</u> (OAR, HR SCHOOL, ETC)	5
<u>OTHER ORGANIZATIONS</u> (TAC, SAC, ETC.)	9

Table 2
SAMPLE DISTRIBUTION: GRADE BY DUTY AFSC

<u>AFSC</u>	<u>CMSGT</u>	<u>2LT</u>	<u>1LT</u>	<u>CAPT</u>	<u>MAJ</u>	<u>LTC</u>	<u>COL</u>	<u>TOTAL</u>
2671	0	14	8	13	2	2	0	39
2675	0	2	16	29	22	5	0	74
261X	0	0	0	1	0	2	0	3
0940	0	0	0	2	2	2	0	6
OTHER*	0	1	0	4	1	2	3	11
73200	1	0	0	0	0	0	0	1
	—	—	—	—	—	—	—	—
TOTAL	1	17	24	49	27	13	3	134

*INCLUDES 0026, 0076, 7516, 7524, ETC.

Job group summaries of incumbent responses to a standard set of job satisfaction questions are shown in Tables 3-5. As can be seen in Table 3, Senior Staff Officers, students, Research Programs Scientists, and Occupational Analysts find their jobs interesting, while a sizeable proportion of the HFE psychologists and the WAPS Test Developers find their jobs dull. As shown in Table 4, most behavioral scientists feel their talents are used fairly well or better; only the Research Task Scientists and WAPS Test Developers have more than 30 percent feeling their talents are not used.

Members of most job groups feel a sense of accomplishment from their jobs as can be seen in Table 5. Only the Senior Staff Officers and WAPS Test Developers have substantial proportions of the group who are dissatisfied or neutral. The Senior Staff officer group is very small (N=5), so that the attitude of just one or two individuals have a great impact on the percent reported.

In an attempt to put these findings into context, data for behavioral scientists were contrasted with a more general Air Force-wide sample. The comparative data is a sample of 2,529 officers in all utilization fields who were part of a study of Professional Military Education (Barucký 1980). They represent most career areas and were asked some of the same basic job satisfaction questions. When compared to the AF behavioral scientists on the question of job interest, the results were mixed (see Table 6): Captains and Lieutenant Colonels who were behavioral scientists reported their jobs interesting in slightly higher proportions than the comparative AF-wide sample. But behavioral scientist Lieutenants and Majors had the opposite trend, with 21 percent of both Lieutenants and Majors finding their jobs dull. In response to the question, "How does your job utilize your talents?" more behavioral scientists responded positively, except for Lieutenants, where 31 percent reported their job used their talents very little or not at all (versus 16 percent of Lieutenants AF-wide; see Table 7).

Finally, in terms of a sense of accomplishment from the job, again most behavioral scientists responded positively, except for the most junior group. The Lieutenants were fairly close to the comparative group (72 percent satisfied versus 77 percent of the AF-wide sample) but a higher proportion (26 versus 18 percent) reported being dissatisfied (see Table 8).

Generally these data present a picture of most Air Force behavioral scientists being more interested and more satisfied with the jobs than their Air Force peers. The exception to this generalization is the most junior behavioral scientists (2nd and 1st Lieutenants) who tend to be dissatisfied to a slightly higher degree than their peers. When this information is combined with the data reported earlier for job groups, it can be recognized that there are several types of behavioral science work where morale is a problem; these include the Human Factors Engineering researchers and the WAPS Test Development psychologists.

Among the WAPS Test Development Psychologists, there are some variations in job satisfaction. An attempt was made to identify tasks or groups of tasks which would identify those Test Developers who are interested and satisfied with their jobs. No conclusive data could be developed, although there was a trend for those who were engaged in off-duty graduate education programs to report job satisfaction than for those not enrolled (those who already have masters degrees).

Table 3
JOB INTEREST OF USAF BEHAVIORAL SCIENTISTS
(PERCENT RESPONDING)

	<u>"MY JOB IS..."</u>		
	(1-3) <u>DULL</u>	(4) <u>SO-SO</u>	(5-7) <u>INTERESTING</u>
RESEARCH PROGRAM SCIENTISTS (N=57)	7	2	91
SENIOR STAFF OFFICERS (N=5)	0	0	100
JUNIOR STAFF OFFICERS (N=9)	0	11	89
ACADEMIC INSTRUCTORS & COUNSELORS (N=16)	6	6	88
RESEARCH TASK SCIENTISTS (N=6)	0	17	83
STUDENTS (N=4)	0	0	100
HUMAN FACTORS ENGINEER PSYCHOLOGISTS (N=9)	22	0	78
OCCUPATIONAL ANALYSTS (N=25)	4	0	96
WAPS TEST DEVELOPERS (N=20)	35	15	50

Table 4
PERCEIVED USE OF TALENTS
(PERCENT RESPONDING)

	<u>"MY TALENTS ARE USED..."</u>		
	(1-2) <u>NOT AT ALL OR VERY LITTLE</u>	(3-5) <u>FAIRLY - VERY WELL</u>	(6-7) <u>EXCELLENT OR PERFECT</u>
RESEARCH PROGRAM SCIENTISTS	7	46	47
SENIOR STAFF OFFICERS	0	100	0
JUNIOR STAFF OFFICERS	11	78	11
ACADEMIC INSTRUCTORS & COUNSELORS	6	44	50
RESEARCH TASK SCIENTISTS	33	51	17
STUDENTS	0	25	75
HUMAN FACTORS ENG PSYCHOLOGISTS	11	78	11
OCCUPATIONAL ANALYSTS	8	76	16
WAPS TEST DEVELOPERS	50	50	0

Table 5
SENSE OF ACCOMPLISHMENT FROM THE JOB
(PERCENT RESPONDING)

	<u>"HOW SATISFIED WITH JOB ACCOMPLISHMENTS..."</u>		
	(1-3)	(4)	(5-7)
	<u>DISSATISFIED</u>	<u>NEUTRAL</u>	<u>SATISFIED</u>
RESEARCH PROGRAM SCIENTISTS	11	2	87
SENIOR STAFF OFFICERS	20	20	60
JUNIOR STAFF OFFICERS	0	11	89
ACADEMIC INSTRUCTORS & COUNSELORS	19	0	81
RESEARCH TASK SCIENTISTS	0	0	100
STUDENTS	0	0	100
HFE PSYCHOLOGISTS	11	0	89
OCCUPATIONAL ANALYSTS	4	8	88
WAPS TEST DEVELOPERS	45	5	50

Table 6
HOW DO YOU FIND YOUR JOB?
(PERCENT RESPONDING)

	<u>LTS</u>		<u>CAPTS</u>		<u>MAJORS</u>		<u>LTC</u>	
	<u>267X</u>	<u>AF SAMPLE*</u>	<u>267X</u>	<u>AF SAMPLE*</u>	<u>267X</u>	<u>AF SAMPLE*</u>	<u>267X</u>	<u>AF SAMPLE*</u>
DULL	21	6	8	6	21	8	0	8
SO-SO	7	9	4	8	0	8	7	7
INTERESTING	72	85	88	86	79	84	93	85

* AIR FORCE SAMPLE OF 2,529 OFFICERS IN ALL AF UTILIZATION FIELDS COLLECTED IN A STUDY OF PROFESSIONAL MILITARY EDUCATION (BARUCKY 1980).

Table 7
HOW DOES YOUR JOB UTILIZE YOUR TALENTS?
(PERCENT RESPONDING)

	<u>LTS</u>		<u>CAPTS</u>		<u>MAJORS</u>		<u>LTC</u>	
	<u>AF</u>		<u>AF</u>		<u>AF</u>		<u>AF</u>	
	<u>267X</u>	<u>SAMPLE*</u>	<u>267X</u>	<u>SAMPLE*</u>	<u>267X</u>	<u>SAMPLE*</u>	<u>267X</u>	<u>SAMPLE*</u>
NOT AT ALL OR VERY LITTLE	31	16	16	12	10	12	0	11
FAIRLY WELL TO VERY WELL	57	62	51	61	41	51	57	51
EXCELLENTLY TO PERFECTLY	12	22	33	27	48	34	43	37

* AIR FORCE SAMPLE OF 2,529 OFFICERS IN ALL AF UTILIZATION FIELDS COLLECTED IN A STUDY OF PROFESSIONAL MILITARY EDUCATION (BARUCKY 1980).

Table 8
SENSE OF ACCOMPLISHMENT FROM YOUR JOB?
(PERCENT RESPONDING)

	<u>LTS</u>		<u>CAPTS</u>		<u>MAJORS</u>		<u>LTC</u>	
	<u>AF</u>		<u>AF</u>		<u>AF</u>		<u>AF</u>	
	<u>267X</u>	<u>SAMPLE*</u>	<u>267X</u>	<u>SAMPLE*</u>	<u>267X</u>	<u>SAMPLE*</u>	<u>267X</u>	<u>SAMPLE*</u>
DISSATISFIED	26	18	14	20	14	19	7	18
NEUTRAL	2	5	2	5	7	5	7	4
SATISFIED	72	77	84	72	79	74	86	77

* AIR FORCE SAMPLE OF 2,529 OFFICERS IN ALL AF UTILIZATION FIELDS COLLECTED IN A STUDY OF PROFESSIONAL MILITARY EDUCATION (BARUCKY 1980).

Each of these major clusters is composed of more discrete job types, which represent variations of emphasis or concentration on some subset of tasks. For example, the Research Programs Scientist cluster includes the following job types: Personnel Research Program Managers, Technology Applications Researchers, Plans Staff Officers, Senior Academic Staff Officers, Contract Monitors, Test Development Researchers, AFMPC Attitude Researchers, and Air War College Evaluators. As might be surmised from the various titles, each group appears to specialize in research activities involving a different organizational mission or program and they perform a slightly different set of specific tasks.

The Academic Instructor-Counselor cluster is composed of four job types: USAFA Instructors-Counselors; AFROTC Instructors-Counselors; USAFA Instructors; and Other Instructors. As these names imply, some USAFA faculty members perform primarily as classroom instructors. A separate group both instructs and serves as counselors. The AFROTC Instructors included in this study are more similar to the USAFA Instructors-Counselors than to the pure Instructor group. Their core tasks reflect considerable personal counseling outside of the classroom. The Other Instructor group includes several unique, one-of-a-kind faculty positions with the DOD Human Relations Institute, AFIT, or other academic units. This group seem to be related in terms of their shared instructing, counseling, and course development tasks.

Most of the other major job groups involved highly specialized functional programs. These include the Occupational Analysts group which is composed mainly of members of the Air Force occupational analysis program at the USAF Occupational Measurement Center. Interestingly, the present USAF officer on exchange duty with the Royal Australian Air Force (in Melbourne, Australia) and the recent USAF-Canadian Exchange position appeared within the occupational analyst job type; even though the positions are located thousands of miles away, the incumbent's response to the job inventory reflected very similar jobs. Within the occupational analyst cluster, three specialized jobs corresponding to sections within the occupational analysis program was identified: inventory development, airman analysis, and officer and management applications jobs.

The Human Factors Engineering Researchers formed a very discrete job group with very little overlap with other groups (although their supervisors and staff personnel did appear in the Research Program Scientist Cluster). Members of this group represent very different organizations including ASD, ESD, the Flight Test Center, and others. Their core tasks reflect a concentration on developmental research efforts (6.3 and 6.4 research) not shared by any other group.

Finally, the Weighted Airman Promotion System (WAPS) Test Development Psychologists assigned to the USAFOMC also appear to have very distinct jobs from the remainder of the career field; two job types within this cluster include test developers and test review psychologists. Their jobs focus almost entirely on procedures for developing Specialty Knowledge Tests and the quality control of such tests. WAPS Test Development Researchers and Section Supervisors appeared in the Research Program Scientist Cluster reflecting the overlap of their jobs with other 267X positions.

CONCLUSIONS

In general, the findings of this study seem to refute the allegations of Jacoby (1970) and other researchers who feel that military psychologists are very dissatisfied and not properly utilized. For Captains and above, our findings reveal Air Force psychologists to be somewhat more satisfied than their Air Force peers. A recent study by Rosenbach and Gregory (1981) of Division 19 members, using Hackman and Oldham's Job Diagnostic Survey (JDS) revealed that military psychologists scored well above national normals of professionals, on most JDS dimensions, except for satisfaction with supervision. Our results suggest that what dissatisfaction there is tends to be focused at the very junior officer level, and may be largely restricted to a few types of jobs. The implication of these findings is that further analysis efforts or management actions should be focused on those job types where a problem may exist, rather than on the entire career field. We should begin to think of Air Force Behavioral Scientists, overall, as being an interested, well-motivated group of professionals.

Our analysis of job types within the behavioral science field suggests that there are a number of very diverse clusters and job types. The implication of this findings is that a return to a four shredout configuration for the AFSC is not practical. To create new shredouts for each major type of job would result in very small populations, and in some cases would effectively limit reassignment actions altogether. Thus, we cannot recommend reshredding the AFSC.

References

- Baruck, J. M., Officer Professional Military Education Curriculum Validation Project. Occupational Survey Report, Occupational Analysis Program, USAF Occupational Measurement Center, Randolph AFB TX, 1980.
- Hedlund, J. L., Newman, S. H., Rasmussen, J. E., Ross, S., Rubenstein, E. A., Sells, S. B., Sperling, P., and Trumbull, R. Military psychology: Comparative Image. (Report of the ad hoc committee on career status of military psychologists, Division 19). American Psychologist, 1968, 23, 112-121.
- Jacoby, Jacob., The Plight of the Uniformed Air Force Psychologist. Professional Psychology, Summer 1970, 1(4), 383-387.
- Rosenbach, W. E., and Gregory, R. A., Job Attitudes of APA Military Psychology Division Members. Paper presented at the 1981 American Psychological Association Convention, Los Angeles, CA, August 1981.
- Phalen, W. J., and Weismuller, J. J., CODAP: Some new techniques to improve job-type identification and definition. Paper presented at the 23rd Annual Military Testing Association Conference, Washington DC, October 1981.

The Ability to Say 'No' as a Mental Health Factor

Allan H. Frankle

College of Education

Drake University

Abstract

The author postulated that the ability to say 'no' is a facet of ego strength associated with mental health, but that arbitrary or deviant nay-saying reflects irrational assertiveness more characteristic of psychiatric patients. Using Lanyon's PSI, a 130 item True-False personality inventory, the hypotheses were tested that normal groups would give more total F but fewer deviant F than psychiatric patient groups. Independent sets of 180 male and 180 female subjects were tested, divided into six subsets of 30 each (normals, neurotics, personality disorders, sociopaths, schizophrenics and organics). Results tended to confirm the hypotheses on both sets: psychiatric patients with diverse diagnoses shared a common tendency to be more submissive and/or more irrationally rebellious than normals. The ratio of 4th root of deviant F to total F synergistically combines the quality and quantity of a subject's ability to resist external coercion or the pressure of his own impulses, differentiating normals at the .001 level from each separate psychiatric subgroup tested. Implications for the armed forces are discussed.

The Ability to Say 'No' as a Mental Health Factor

The ability to say no is a fundamental act asserting one's autonomy and individuality. Those who are excessively dependent on others, who are intimidated by authority, who are incapable of critical thinking, who have never internalized a set of values, find it difficult to say 'no'.

The tendency toward suggestibility and docility of the mentally retarded has been recognized for centuries. From a political science viewpoint John Stuart Mill (1859, 1963) in his essay, *On Liberty*, wrote: "...a State which dwarfs its men, in order that they may be more docile instruments in its hands even for beneficial purposes -- will find that with small men no great thing can really be accomplished."

As ego strength and the sense of identity and autonomy grow, the individual acquires the power of self-discipline as well as the power to resist external coercion. We all have to learn to say 'no' to ourselves, to our own impulses, at times. The literary expression, "a slave to his desires," is not so far-fetched; and the widespread incidence of various addictions is testimony to the difficulty of saying 'no' to oneself. The popular concept of "won't power" as the obverse side of "will power" embodies this idea.

In the course of growth toward autonomy most persons go through a period of struggle during which they swing episodically toward excessive or irrational opposition. Some even become locked into a chronic negative attitude, where the principle of saying 'no,' of disagreeing and disobeying, becomes the dominant pattern of life, without regard to the circumstances that prevail. This can be recognized as irrational defiance or passive aggression.

Ultimately, the person who reaches a state of emotional and social maturity can say 'no' and stand firm in his position when the facts and circumstances warrant such a stance; yet he retains the flexibility and rationality to say 'yes' when agreement is appropriate.

A different, more cognitive interpretation of the foregoing issue is that a rational negative response can be a manifestation of critical thinking, as opposed to fuzzy or overinclusive thinking. The computer science aphorism, "garbage in garbage out," undoubtedly has relevance for mental health. Persons who think more precisely about the semantic validity of the statements they hear and read filter out a sizable proportion of the semantic "garbage" that comes their way, hence may have better-functioning computers. This interpretation seems congruent with Ellis's (1977) Rational-Emotive concept of mental health.

The somewhat different interpretations of the ability to say 'no' as a manifestation of will or ego-strength versus critical thinking need not be irreconcilable. Both factors may be operating simultaneously in varying degrees, and one could view critical thinking as another facet of ego strength.

Bearing in mind the background just covered, we may now ask: To what extent can we measure the quality and quantity of assertiveness by the use of a personality inventory, and is it related to overall mental health? The last part of this question can be tested by comparing inventory scores for normals with various psychiatric patient groups.

Method

Subjects: Two major sets of subjects, one male and one female were utilized to test the hypotheses of the research. Each set was divided into six subsets of approximately 30 subjects each. The subsets include well-adjusted normals, organics (neurologically injured or ill), schizophrenics, sociopaths, personality disorders and neurotics. The patients consisted of persons who were outpatients or inpatients of a general hospital psychiatry service or were treated by the author as private outpatients. All were studied in sufficient depth to establish a differential diagnosis. All were capable of reading the PSI adequately and, as far as could be determined, were willing to cooperate on the testing procedure used.

The patients ranged in age from about 13 through 65, with most falling in the young adult to middle-age bracket. The organic patients showed a mean in the early 30s, and included very few senescent individuals. The female normals were mainly student nurses in training in three hospital schools of nursing, while the males were mainly volunteers who showed a history of reasonably good adjustment. The normals averaged about 2½ years of college education, while the mean of the psychiatric groups was about high school graduation. Thus the normals were somewhat younger, brighter and better educated than the psychiatric patients, but not greatly so.

Materials: The only instrument used for collecting data was the Psychological Screening Inventory (Lanyon, 1973), a 130-item True-False personality questionnaire. This test is printed on both sides of a single 8½" by 11" sheet. Subjects mark their answers directly on the test sheet. It has been found that the task is made easier and more fool-proof by not requiring a separate answer sheet, and that the brevity of the test (only 15 to 20 minutes) is also advantageous when working with patients who are uneasy, restless, or have a short attention-span.

Procedure: The test was administered by a clinical psychologist or a psychometrist trained in the use of the PSI. The basic instructions were to answer all the questions as honestly as possible, that there was no time limit, that if a statement tended to be true it was to be marked True, and vice-versa. A few omissions were permitted, but patients whose tests were not completed, who were clearly uncooperative, who could not read and understand the first five items, or who answered all True or all False, etc. were not included. Such cases constituted only a very small percentage of the tests administered, as most of the patients were in adequate contact, cooperative, and had more than an eighth grade education. About half of the PSI items are written so that the normal answer is True and about half False. This ratio could not be maintained for the Alienation scale, however, and on this measure of psychoticism the normal answer is more often False.

It is proposed that the total number of F responses to the PSI be treated as a metric for measuring self-assertion. Low F reflects submissiveness and inadequacy, and by inference, poor mental health. This total number of F responses may then be partitioned into consensual F and deviant F by the use of Lanyon's table (1973, p. 33) showing the proportions, by sex, of 400 normal males and 400 normal females who answer True to each item. A cut-off point of proportion True = .687 or higher for a given item was chosen to classify a False response to that item as deviant. A count was made of deviant F responses

for all subjects, using special stencils made up for male subjects and female subjects. Because of significant sex difference in p values, separate scales are needed for males and females. This number was hypothesized as representing irrational self-assertion. It is inferred that larger deviant F scores are undesirable from a mental health standpoint. No subjects participating in this research played any part in the construction of the Deviant F scoring keys.

Psychometrically we have here an interesting situation: Higher total F is postulated as a positive indicator of mental health, but higher deviant F is regarded as a negative indicator of mental health. It is therefore proposed that a ratio be utilized between the number of deviant F responses (numerator) and the number of total F (denominator), which would yield a quotient to serve as an index of maladjustment or illness. High quotients would, according to the hypotheses just stated, reflect an excess of irrational negativism or passive aggression demonstrated by too many deviant F responses, and/or a general lack of assertion (passivity, timidity, submissiveness) shown by too few overall F responses. It was found that in order to synchronize the variances of the numerator and denominator, the optimum formula was actually the fourth root of deviant F divided by total F. The concept of a "moderator variable" might be kept in mind when considering this ratio.

A t-test of the differences between the mean total F, mean deviant F and mean ratio was computed, comparing normal males with each of five male psychiatric patient groups, and this same procedure was followed for the female subjects as an independent replication.

Results

The findings related to the hypotheses can be read most clearly from a few tables. The first table presents the total number of False responses to the 130 items of the PSI by the two major groups (12 subsets) showing the means, S.D.'s, t-ratios of means of patient groups vs. normals, and number and percent of each subset cut by optimum critical score.

Enter Table 1 about here

The second table presents just the number of deviant False responses to the PSI for the two major groups, by subsets, showing means, S.D.'s, t-ratio vs. normals, number and percent cut by optimum cutting score.

Enter Table 2 about here

The third table presents the subset means of the two major groups for the fourth ratio: fourth root of the number of deviant F responses, divided by total F. A maximum score of 70 for the total F of the denominator was used because it was found that F scores larger than 70 did not reflect better

Table 3

Analysis of Ratio: $(4\text{th Root of Deviant } F \div \text{Total } F^a) \times 10^3$

by Subsets

Male Subsets	N	Mean	S. D.	t vs Normals	Cut by c.s. ≥ 31.3	
					#	%
Normals	30	26.35	3.224	-----	2	7
Neurotics	30	30.22	3.741	4.22***	11	37
Pers. Disorders	30	30.44	4.999	3.42**	12	40
Sociopaths	30	30.57	3.124	5.06***	12	40
Schizophrenics	30	33.92	4.635	7.22***	22	73
Organics	30	34.56	4.938	7.49***	20	67

Female Subsets	N	Mean	S. D.	t vs Normals	Cut by c.s. ≥ 27.8	
					#	%
Normals	30	23.75	2.532	-----	0	0
Neurotics	30	28.83	4.161	5.62***	17	57
Pers. Disorders	30	29.27	3.605	6.74***	18	60
Sociopaths	30	30.02	3.514	7.79***	21	70
Schizophrenics	30	32.45	5.625	7.59***	24	80
Organics	30	33.79	6.512	7.73***	25	83

^a Maximum value for total F set at 70

** = p .01

*** = p .001

Table 2

Analysis of Deviant "False" Responses by Subsets

Male Subsets	N	Mean	S. D.	t vs Normals	Cut by c.s. ≥ 18	
					#	%
Normals	30	8.87	4.256	----	2	7
Neurotics	30	12.63	4.475	3.28**	3	10
Pers. Disorders	30	11.83	3.724	2.82**	2	7
Sociopaths	30	12.77	5.226	3.12**	6	20
Schizophrenics	30	13.90	4.621	4.41***	7	23
Organics	30	10.53	3.774	1.58	1	3
Female Subsets					Cut by c.s. ≥ 13	
					#	%
Normals	30	5.90	2.890	----	1	3
Neurotics	30	9.90	4.213	4.21***	6	20
Pers. Disorders	30	10.13	3.373	5.13***	8	27
Sociopaths	30	12.40	4.909	6.14***	11	37
Schizophrenics	30	12.83	4.517	6.96***	17	57
Organics	30	9.06	3.915	3.50***	3	10

** = $p < .01$ *** = $p < .001$

Table 2

Analysis of Deviant "False" Responses by Subsets

Male Subsets	N	Mean	S. D.	t vs Normals	Cut by c.s. ≥ 18	
					#	%
Normals	30	8.87	4.256	----	2	7
Neurotics	30	12.63	4.475	3.28**	3	10
Pers. Disorders	30	11.83	3.724	2.82**	2	7
Sociopaths	30	12.77	5.226	3.12**	6	20
Schizophrenics	30	13.90	4.621	4.41***	7	23
Organics	30	10.53	3.774	1.58	1	3
Female Subsets					Cut by c.s. ≥ 13	
					#	%
Normals	30	5.90	2.890	----	1	3
Neurotics	30	9.90	4.213	4.21***	6	20
Pers. Disorders	30	10.13	3.373	5.13***	8	27
Sociopaths	30	12.40	4.909	6.14***	11	37
Schizophrenics	30	12.83	4.517	6.96***	17	57
Organics	30	9.06	3.915	3.50***	3	10

** = $p < .01$ *** = $p < .001$

adjustment. Total F scores greater than 70 were entered in the denominator as 70.

Enter Table 3 about here

Discussion

Basically, results for 180 subjects in the male group and the replication of the tests on the female group of 180 support the hypotheses proposed. The array of means for total F shown in Table 1 for the 12 subsets, though running in the right direction, shows significant differences only when comparing the normals to the schizophrenics and organics. The normals do show a higher mean level of total F than any other subset in both the male and female groups, from which it is inferred that normals are, overall, more able to assert themselves than are patients.

The data for means of the deviant F responses show somewhat stronger differentiation of normals from most of the clinical groups than did total F, doing least well on the organics, who were especially well distinguished by their low total F. It seems quite clear that psychiatric patient groups tend to give more deviant F responses than normals, from which it is inferred that they are prone to irrational opposition or negativism at times.

Of special interest is the synergistic combination of these two related variables, best displayed by the ratio of the 4th root of the deviant F to total F. The 4th root of deviant F was employed because this best synchronized the variances of the numerator and denominator. This ratio provided the best discrimination between the normal group and all clinical groups, with higher scores associated with maladjustment or mental illness. Examining not only the t-test data but the percentage of each group exceeding the critical scores set for the male and female groups, we can see that a single screening score cuts a relatively small proportion of our well-adjusted normals (i.e., a low base rate of false positives among normals), while cutting a rather substantial proportion of the various psychiatric patient groups. The ratio is clearly superior to either component separately, and differentiates normals at the .001 level from all psychiatric sub-groups tested. It is to be noted that the scores shown contain no correction for defensiveness. When a correction is applied it brings about still better differentiation between normals and patients, but this point will not be elaborated here as it is not essential to the basic theory.

These findings are of special interest because they reveal a common element in the behavior of patients who bear very different diagnoses. The nay-saying set of an individual may legitimately be considered a "behavior sample" whether on a psychological test or in everyday life. It may seem difficult to reconcile inadequate nay-saying as a deficit of the power of self-assertion with the tendency toward irrational or deviant F responses. But Rollo May in his book, Power and Innocence (1972), makes the point that a sense of helplessness or lack of control of one's own destiny can ultimately generate rage. If total

F reflects the sense of autonomy or power, and deviant F contains an element of rebellion, it seems plausible that there can be some kind of interaction between these variables even in the process of responding to the PSI.

The useful interaction between the 4th root of deviant F and total F in a ratio relationship is congruent with the concept of a moderator variable. In this instance it would appear the deviant F is the primary predictor, most closely related to the subject's classification as normal or patient. But the positioning of total F, a variable less closely correlated with the criterion, in the denominator markedly enhances the relationship. Some peripheral research not yet completed indicates that total F can serve effectively as a moderator variable for some other PSI scales, including the already admirable Alienation scale developed by Lanyon, yielding modest incremental validity.

The military services provide a special crucible in which officers and enlisted personnel can be heavily stressed. Although the stress is probably greatest during times of war, peace-time service does have its psychological pressures, and among the prominent pressure situations are officer-training programs at the service academies and OCS centers. A special problem may be generated in the form of internal conflict between the demand for absolute obedience to command authority and the expectation that an officer will show fierce and determined resistance to the enemy, as well as demonstrating decisiveness and firm resolve as a leader of his own unit.* To handle such seemingly dissonant roles (assertion and obedience) requires considerable flexibility as well as sound judgment of situations and people. It would be of interest to see whether the scores referred to in this paper have any predictive validity for performance of cadets during their officer-training; and also to see whether their scores on such an instrument will show significant changes from year to year as a result of the type of training they have undergone, using a matched sample of students in conventional colleges of universities (not in ROTC) as controls.

In recent years more attention has been devoted to the concept of life stress, built on basic research such as Selye's (1976) and extended by Holmes and Rahe (1967) to take into consideration life events as psychological determinants of illness. Some persons fall ill under the impact of only small doses of stressful events, while others thrive on heavier doses. The 4th root ratio may help to measure "resistance to stress," which could have value for selective placement of military personnel.

Finally, it should be noted that the concepts and methods put forward in this paper may not be limited to application via the PSI. Other personality inventories could serve equally well or even better if they are reasonably subtle and constructed so as to pull roughly equal numbers of true and false answers significant for psychopathology. Much information on both normals and patients is undoubtedly available on file, which could be analyzed retrospectively to verify or refute the hypotheses of this research.

References

- Ellis, A. & Grieger, R. Handbook of rational-emotive therapy. New York: Springer Publishing Company, 1977.
- Holmes, T. H. & Rahe, R. H. The social readjustment rating scales. J. Psychosom. Res., 11: 213-218, 1967.
- Lanyon, R. Psychological Screening Inventory: Manual. Goshen, New York. Research Psychologists Press, 1973.
- May, R. Power and Innocence. New York: Norton and Company, 1972.
- Mill, J. S. On Liberty, Representative Government, and The Subject of Women-Three Essays. London: Oxford University Press, 1963.
- Selye, H. The Stress of Life. New York: McGraw-Hill, 1976.

THE AFMET PROJECT AND THE BLOOM SENTENCE COMPLETION SURVEYS

Wallace Bloom, Ph.D, Wilford Hall USAF Medical Center

INTRODUCTION

Unresolved psychological problems that existed prior to enlistment in the USAF often contributed or led to poor emotional adaptation to military life and then to involuntary separation. The various services have researched ways to identify in advance, the individuals with significant psychological problems who would fail to function effectively in the military environment (Bucky and Edwards 1974, Guinn, Johnson and Kantor 1974, LaChar et al. 1974, Plag 1962). The Air Force Medical Evaluation Test (AFMET) was begun in June 1975 (Bloom 1977, 1980, O'Hearn and Bloom 1978). In Phase I, each airman entering basic training at Lackland AFB was given a TRUE/FALSE history-opinion-interest inventory (HOI) and 457,467 had been tested by 30 September 1981. Those identified by computer scoring as low mental health risks (approximately 93.3%) continued basic training without further mental health screening. In Phase II, those not identified as low-risk 30,458 were called in for individual mental status interviews (Bloom 1981) with a mental health technician at their squadrons and additional psychological tests were administered. The reports of these interviews and tests were reviewed by a clinician who determined which additional trainees (approximately three out of four) might then be identified as low risk and not require further evaluations. In Phase III, those individuals not already identified as low risk are referred to the Mental Hygiene Clinic for clinical assessments and appropriate followup actions i.e. return to duty, referred to other agencies, or diagnosed and recommended for discharge consideration. To date 2,121 were recommended for discharge, and 2,634 were discharged by other agencies before AFMET evaluations were completed.

BSCS

The Bloom Sentence Completion Survey (BSCS) (Bloom 1975, 1979, 1980) has been used since 1 January 1976 in AFMET Phase II. The test items had been constructed and selected to reveal each subject attitudes toward seven (7) areas: People, Physical, Self, Family, Psychological Self, Self-Directedness, Work, and Accomplishment. After five hours of instruction and several weeks of supervised practice, it was rare that the scoring of the enlisted mental health technicians differed on more than two (2) of the thirty-five (35) scored items from that of the instructor. Inter-rater reliability was over 94 percent.

The MMPI initially used as the Phase II test was found to be not suitable as a screening device even after new Air Force norms were established (Bloom 1977 b). However, it is now used in Phase III clinical assessment on an "as needed" basis.

Validity

The validity of this sentence completion survey was tested against the criterion of retention or discharge. Table 1 illustrates that during the year ending February 1, 1979, there were 3,541 adults tested and each and every subscore separately was significantly related to outcome. Noteworthy was that the composite score of those returned to duty (RTD) after the Phase II interviews were more than 14 points higher (9.3771) than those discharge after Phase II (-5.4724).

Stability

The Stability of the test over time was illustrated in Table 2. During five separate increments of 19,712 subjects tested between January 1, 1976 and October 1, 1981 there were general similarities in average composite scores and the subscores showed only slight differences over the six years. During these years there were 42 technicians evaluators scoring these tests, and none were present the entire time. Usually there were only 11 assigned at any one particular time. Properly trained technicians came up with reliable scores year after year

Excessive Neutral Responses

When an analysis sheet shows four or more neutral scores, it has proven helpful to go back and review these items. Three typical characteristics have been identified. First, there are the clients who are evasive, such as one who indicated (Item 10) MY BODY IS my body. (Item 2) WHEN I LOOK AT MYSELF, I see myself. Secondly are those who hedge or are noncommittal by combining both positive and negative statements in one response. Equivocators have responded: "WHEN I LOOK AT MYSELF I SEE some things I like and some I don't like... WHEN I'M UNDER PRESSURE OR THE WORK GETS HARDER, I work harder or quit." Thirdly are those who do not complete a test item even after inquiry or help from the examiner. These are avoiders and careful interviewing often can reveal exactly what each is avoiding and the experiences of feeling involved. A second look at the neutrally scored test items has helped to identify defensive areas of EVADERS, EQUIVOCATORS and AVOIDERS and facilitated further clinical assessments.

Suicidal Ideations

Early in this project, it was observed that some responses to BSCS Item 12, often suggested extreme depression and/or suicidal ideas. Examples are:

THERE ARE TIMES WHEN I:

Want to crawl under a rock and die.
Wish I were dead.
Feel like running away.
Wonder why I was ever born.
Could shoot myself.

There also were some individuals, so preoccupied with thoughts of suicide, that mentally they could not project themselves into the future. We called this a "Brick Wall Factor." Stem items originally selected to elicit attitudes toward ACCOMPLISHMENT because they focused on future activities, at times were answered in a grossly negative way or not answered. For example:

Item 15. TEN YEARS FROM NOW:

I don't know
I may be dead.
If I will be alive.
Want to be Dead.
Probably won't be living.

Item 31. LOOKING AHEAD TOWARD THE FUTURE.

I may not have a future.
See nothing for me.
I doubt if I'll amount to anything.
I just live from day to day.

Item 39. SOMEDAY I WILL:

Probably kill myself.
Die.
That's a stupid question
Won't have anything to worry about. (?) be dead.
Go too far. (?) Get hurt, (Patient drives recklessly).

Suicidal thoughts also have been suggested by responses to other stem items.
For example:

- Item 20. SECRETLY, I wish I were dead.
Item 4. MOST PEOPLE DON'T KNOW THAT I think about death.
Item 36. WHEN I LET MYSELF GO I do reckless things. (Tell me more.) I drive too fast, once at night I drove on the wrong side of the freeway, just for kicks.
Item 28. PEOPLE THINK OF ME AS not doing what I say, but someday I'll show them.

Indications of more than normal suicidal risks can and have been noted in at various times, in almost any test item. These are not limited to items 12, 15, 31, and 39. Trainees whose responses indicated questionable suicidal risks were referred to Phase III or placed in protective hospitalization. In the six years of the project and the screening of almost half a million trainees there have been no suicides although approximately 15 would have been expected based on National averages.

Conclusion

In the first six years of the AFMET Project over 450,000 trainees were screened and more than 30,000 individually interviewed and tested. Responses to the Bloom Sentence Completion Surveys clinically helped us to interview, reveal defensive coping styles, evaluate suicidal risks, and statistically suggest possible discharges. A data bank of demographic and mental health information has been established and is to be further analyzed to identify factors relating to success or failure in military service

REFERENCES

- Bloom, W. Manual, Bloom Sentence Completion Surveys; Student and Adult Effectiveness Training Associates of San Antonio, 133 Twinleaf Lane, San Antonio, Texas 78213, 1975
- Bloom, W. Air Force Medical Evaluation Test, USAF Medical Service Digest 28,(2) 1977a.
- Bloom, W. Relevant MMPI Norms for Young Adult Air Force Trainees. Journal of Personality Assessment 1977b 41(5) 505-510.
- Bloom, W. Manual: Bloom Sentence Completion Survey: Student, Chicago, Illinois Stoelting Co. 1979.
- Bloom, W. Manual: Bloom Sentence Completion Survey: Adult, Chicago, Illinois Stoelting Co. 1980a.
- Bloom, W. Air Force Medical Evaluation Test (AFMET) Identifies Psychological Problems Early. USAF Medical Service Digest Nov-Dec 1980b, 8-9.
- Bloom, W. Assessment of Patients: The Standardized Report of Interview (SRI). Military Medicine, 1981, 146(9) 639-641.
- Bucky, S. F. & Edwards, D. The Recruit Temperament Survey (RTS) as it discriminates between psychoses neuroses, and personality disorders. Journal of Clinical Psychology 1974, 30, 195-199.
- Guinn, N., Johnson, A. L. & Kantor, J. E. Screening for adaptability to military service. Report no. AFHRL-TR-75-30 Air Force Resources Laboratory, May 1975.
- LaChar, D., Sparks, J. C. Larsen, R. M. , & Bisbee, C. T. Psychometric prediction adaptation for USAF basic trainees. Journal of Community Psychology 1974, 2(3), 268-277.
- O'Hearn, T. P. & Bloom, W. The Air Force Medical Evaluation Test (AFMET) Program: A Three-Phase Assessment Program. USAF Behavioral Sciences Symposium May 24, 1978.
- Plag, J. A. Pre-enlistment variable related to the performance and adjustment of Navy recruits. Journal of Clinical Psychology 1962, 19,168-171.

TABLE 1
BLOOM SENTENCE COMPLETION SURVEY SCORES
USAF BASIC TRAINEES, LACKLAND AFB, TEXAS
February 1, 1978 to February 1, 1979

	ALL		n = 2599		n = 867		t
	n = 3541*		Phase II-RTD		Phase II-MHC		
	\bar{x}	sd	\bar{x}	sd	\bar{x}	sd	
People	-1.0345	2.792	-0.7995	2.795	-1.7070	2.720	8.45**
Physical Self	.8560	2.8371	1.3190	2.694	-0.4302	2.710	16.48**
Family	1.1986	3.0205	1.5975	2.825	0.0358	3.287	12.53**
Psychological Self	-0.0997	2.9780	0.4871	2.896	-1.7785	2.579	21.70**
Self-Directedness	1.8718	2.5805	2.5979	2.199	-0.1788	2.507	29.09**
Work	0.8543	2.6334	1.2301	2.732	-0.2099	2.056	16.36**
Accomplishment	2.3507	3.0085	2.9450	2.746	0.7197	2.831	20.19**
Composite Score	5.9952	11.5312	9.3771	10.018	-3.5490	10.306	32.20**

	n = 688#		n = 199		t
	MHC	RTD	MHC	Discharge	
	\bar{x}	sd	\bar{x}	sd	
People	-1.6192	2.974	-1.6382	2.069	0.10
Physical Self	-0.0843	3.247	-0.6482	2.574	2.56*
Family	0.2747	3.334	-0.1256	2.937	1.64
Psychological Self	-1.3677	3.148	-2.0050	2.251	3.19**
Self Directedness	0.5625	2.653	-0.8040	2.328	7.06**
Work	0.1831	2.172	-0.4523	1.961	3.93**
Accomplishment	1.2631	3.300	0.2010	2.476	4.92**
Composite Score	-0.7878	11.503	-5.4724	10.046	5.60**

NOTE: *Total subjects tested was greater than sum of subgroups as 75 subgroups were not interviewed at Phase II due to intervening circumstances.

#20 subjects who by-passed Phase II or were Incompletely Identified there were seen at Phase III, Mental Hygiene Clinic.

RTD — Return to Duty

MHC — Referred to Mental Hygiene Clinic

BSCS SCORES OF BASIC TRAINEES IN THE AFMET PROGRAM
LACKLAND AFB, TEXAS
January 1, 1976 to February 1, 1979

	A Jan. 1 - Jun. 15, 1976 n=2879*		B Sep. 28, 1976 Jan. 31, 1978 n=3977 Adults		C Feb. 1, 1978 Feb. 1, 1979 n=3041 Adults		D Feb. 1, 1979 Oct. 1, 1980 n=5762		E Oct. 1, 1980-1981 n=4053 Adults	
	\bar{x}	sd	\bar{x}	sd	\bar{x}	sd	\bar{x}	sd	\bar{x}	sd
People	-0.59	1.98	-1.11	2.11	-0.97	2.67	-1.245	2.016	-1.226	2.097
Physical Self	0.65	2.25	.89	2.38	1.03	2.73	1.144	2.368	1.072	2.350
Family	1.25	2.47	1.18	2.60	1.25	2.97	1.174	2.651	1.370	2.670
Psychological Self	0.03	2.17	.04	2.35	0.07	2.97	.053	2.401	.253	2.374
Self-Directedness	2.06	2.13	1.99	2.29	1.91	2.61	1.943	2.354	2.015	2.326
Work	0.93	1.79	1.00	1.85	0.95	2.71	1.161	1.789	1.163	1.709
Accomplishment	2.44	2.15	2.54	2.13	2.37	2.80	2.427	2.166	2.456	2.108
Composite Score	6.74	NA	6.54	NA	6.47	11.51	6.732	9.783	7.115	9.430

*Includes less than 10% Students, as separate records were not available for that period.
Standard Deviations on Composite Scores were not available for the earlier years.

HUMAN DEVELOPMENT CENTER - A MOVEMENT TOWARD PREVENTIVE MENTAL HEALTH

DAVID A. GILES, Ph.D.
CAPT., USAF, BSC
SHEPPARD USAF REGIONAL HOSPITAL
SHEPPARD AFB, TX 76311

A unique set of environmental and interpersonal demands are associated with the first six months of an individual's enlistment in the military service. During basic military training and technical training school, individuals experience developmental crises, such as separating from family and friends, conforming to external demands, and attempting to establish self-identity and autonomy. This developmental stage is usually a time of rapid personal growth and accomplishment; however, adjustment problems can adversely influence the student's adaptation to the military to the degree that administrative separation becomes the only feasible solution.

In 1978, the Air Training Command (ATC) launched the Human Development Center (HDC) concept. This experimental program was designed to meet the unique needs of students in training and reduce the expense of student training failures. The HDC mission at Sheppard AFB was to "save training dollars" by providing early identification and prompt mental health services for the 4500 students assigned to technical training. This goal was achieved by devising an intake system that included a psychiatric technician's assessment, as well as an evaluation by a mental health professional. The time from referral to written response averaged five (5) days (with less time required for a verbal response). Special evening clinic hours were established to facilitate the evaluation process and reduce the student's time spent away from the classroom. Workshops in providing accurate and useful data on the mental health evaluation requests were also conducted with those individuals who frequently refer students.

The scope of the HDC mission rapidly increased to include providing mental health services to the school staff (instructors) and other permanent party assigned to the student training squadrons. In 1980, the five professionals and three technicians assigned to Sheppard's HDC provided direct patient-care services to over 1600 individuals, in addition to establishing preventive approaches such as the student liaison program.

STUDENT SQUADRON LIAISON PROGRAM

Mental health providers accept the responsibility for showing how psychological, emotional, and social factors impact upon the individual, in addition to assessing what interpersonal and environmental supports exist. These goals are usually accomplished by one or more of the following approaches: (a) Tertiary prevention--designed to lower the incidence of emotional problems by providing prompt treatment and after-care services; (b) Secondary prevention--designed to reduce the

incidence of emotional problems by early identification of problems before they become severely dysfunctional; and (c) Primary prevention--designed to lower incidence rates by elimination of harmful social and environmental stress by promoting positive social change within the technical training environment.

Initial efforts by the HDC staff were directed toward providing tertiary and secondary prevention. The student squadron liaison program placed a HDC staff member in each of the five technical training groups. The initial focus was on improving the line officer's (commanders and instructors) ability to identify students with potential emotional problems. Several formal workshops on early identification presented to squadron commanders and staff were later expanded and conducted with technical school instructors. HDC staff members also sponsored an annual technical training center-wide social event involving a "Texas Goat Roast and Chili Cookoff". This social event facilitated the communication between squadron commanders and school staff, as well as promoted HDC's involvement in the training operation.

The adoption and incorporation of HDC's preventive mental health approaches are due in large part to the direct communication efforts of HDS staff members with the line and unit commanders. A computer-assisted review of the 1980 HDC records reflected that a majority of the student referrals had no significant diagnosis or intervention recommendation. This finding indicated a great deal of wasted time was occurring in formulating mental health evaluations that contained few recommendations; consequently, modification of the referral pattern became a prime HDC goal. HDC expanded its operation to two satellite offices within the classroom buildings to screen referrals prior to the initiation of a formal mental health evaluation request. This change not only reduced unnecessary referrals but avoided students being unnecessarily out of the classroom, provided feedback quicker, and provided the opportunity for further HDC program development. The commander's acceptance and implementation of HDC recommendations significantly fostered the evolution and enhancement of mental health services at the Sheppard Technical Training Center.

FUTURE DIRECTIONS

Efforts are presently underway to increase the effectiveness of medical care by interfacing with the preventive case-finding efforts conducted at the basic military training center. The Air Force Medical Evaluation Test (AFMET) is a screening-device employed during the first weeks of basic training to identify students with a "high-risk" for developing adjustment and/or emotional problems. This identified "high-risk" population frequently demonstrates difficulty adapting to the demands of military service both in basic training and advanced technical training. Sheppard's HDC is pioneering efforts to utilize the AFMET program data to facilitate early identification and provide prompt interventions when necessary.

The preventive mental health model at Sheppard AFB, which emphasized extending mental health services beyond the clinic or hospital setting, has found tremendous acceptance with most base organizations. HDC staff members have been able to conduct adjustment seminars at the Correctional Custody Facility (CCF) and Security Police (SP) squadron. Efforts are also in progress to provide stress management techniques to rated personnel of the 80th Flying Training Wing. The guiding principle in devising these consultation approaches has been

to make recommendations as pragmatic and relevant as possible to the consumer's needs. The integration of the mental health professional and the line commander has resulted in an improved health-care delivery system and reduced the counter-productive effects that mental health problems exert on organizational effectiveness.

THE INTERPERSONAL DIMENSIONS OF THE PASSIVE-DEPENDENT AND COMPULSIVE PERSONALITIES IN THE MILITARY

Frank E. Emmett, Jr., Ph.D., Wilford Hall USAF Medical Center

INTRODUCTION

In 1969 Theodore Millon attempted to clarify many of the issues that have plagued clinical psychology's attempts to devise a system of nosology in psychopathology which would be testable and verifiable. He set as the cornerstone of his theoretical framework the concept of the basic personality style. In enumerating eight basic patterns and their accompanying dysfunctional behaviors, Millon set forth the possibility of verification of these behavioral signs by other clinicians.

A great deal of interest in this type of specificity of behaviors in psychopathology has been generated by the publication of the third edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-III) by the American Psychiatric Association (1979). Millon served as senior consultant to DSM-III in the area of personality disorders, and many of his concepts were utilized. In addition, there has been increasing interest in the literature, and within the framework of DSM-III, about the role of personality disorders in adjustment and decompensation. In a manner similar to Millon's, DSM-III has made the personality disorder the keystone of more severe functional disorders.

This study assessed the specified behavioral components of two of Millon's basic personality styles observed within a military outpatient mental health environment. Using the guidelines and concepts set down by Millon, pencil and paper instruments were used to measure the specified behaviors of the two personality styles.

METHODS AND PROCEDURES

Millon's system is based upon the occurrence or nonoccurrence of certain behaviors within the interpersonal domain of the individual. In developing the personality disorder as the keystone of his theory Millon has utilized four areas of functioning which give rise to the most descriptive picture of the individual's behavior:

1. Interpersonal Perception -- how the individual is perceived by others
2. Cognitive Functioning -- what composes the mental set of the individual
3. Self-Image -- how the individual sees himself
4. Interpersonal Behaviors -- how the individual behaves towards others.

Of the eight basic personality disorders specified by Millon, a prior pilot study conducted by the author had isolated the passive-dependent (Style C) and compulsive personality (Style G) as the personality styles most frequently seen in a military clinical population.

The task then became one of matching the behaviors specified by Millon under the four central traits for the two personality styles with specific clinical tests or scales that would measure the presence and degree of that behavior.

Examining Millon's central traits for each personality style under investigation yielded the following descriptions:

I. INTERPERSONAL PERCEPTION

Passive-Dependent

"Gentle Affability -- kind, soft, humanistic, inadequate, responsible to others, obliging, grateful."

Compulsive

"Restrained Affectivity -- emotionally controlled, grim, cheerless, respectful, duty-bound, moralistic."

II. COGNITIVE FUNCTIONING

"Cognitive Denial -- refusing to admit discomforting thoughts about self or others."

"Cognitive Constriction -- narrow-minded, overly methodical and pendent in thought."

III. SELF-IMAGE

"Inadequate -- views self as inferior, fragile, unworthy, cooperative, considerate, unambitious, pitiful."

"Conscientious -- views self as practical, prudent, moralistic, responsible, tense, perfectionistic."

IV. INTERPERSONAL BEHAVIOR

"Interpersonal Compliance -- willing to submit and oblige others, conciliatory, suggestable, placating, helpless, sacrificing, clinging."

"Interpersonal Respectfulness -- ingratiating with superiors, formal and legalistic with subordinates, conforming, polite, deferential, autocratic, ritualistic."

Using the guidelines and concepts set-down by Millon, pencil and paper instruments were selected to measure the specified behaviors of the two personality styles. The instruments involved in this study were: (1) The Sixteen Personality Factor Questionnaire -- 16PF (Cattell, 1968); (2) The California Personality Inventory's Dominance and Femininity Scales (Gough, 1957); (3) Fundamental Interpersonal Relationship Orientation -- Behavior, FIRO-B (Schutz, 1966); (4) Rehfisch Rigidity Scales (Rehfisch, 1958); (5) Rotter Internal/External Scale (Rotter, 1966); and (6) Interpersonal Behavioral Inventory -- IBI (Lorr et al, 1965). Coupling the four central traits and the instruments designed to measure the trait components yielded Table 1.

The research population consisted of 50 persons for each of the two personality styles (25 males/25 females) for a total sample of 100. The subjects were active duty military members who were seeking outpatient psychotherapy or evaluation, and who volunteered to participate in the research. (All research procedures and protocols for the use of subjects were cleared through the Wilford Hall Human Subjects Committee.) Psychologists and social workers at the Mental Hygiene Clinic, WHMC, Lackland AFB, TX served as interviewers and raters who assessed the individual's personality style to determine if he/she possessed one of the styles under investigation. Prior to the start of the research an extensive training program

in Millon's criteria was conducted by the Experimenter. After an in-depth interview if the initial rater believed that the subject possessed one of the personality styles under investigation the subject was referred to a second rater who also interviewed him/her. When there was concurrence between the two raters that the subject possessed the same personality style, the subject was administered the battery of test instruments by the second rater, while the initial rater completed the behavior rating scale (IBI).

RESULTS

The initial computations involved a multivariate 2 (personality style) X 2 (sex) analysis of variance (MANOVA) with the 28 different scales as dependent measures. When the results of any of these MANOVAs were significant, univariate 2x2 analyses of variance (ANOVA) were performed on the scales which served as criteria in the MANOVA. Pillai's trace statistic was utilized to obtain approximate F statistics due to its superior discriminating power and robustness in dealing with diffuse structures (Olsen, 1976).

The results from the MANOVA indicated that there was a significant difference between the Passive-Dependent and Compulsive personalities, as well as between males and females along the 28 dependent variables. The null hypothesis for no interaction between personality style by sex could not be rejected. (Table 2)

Looking at the ANOVAs for the four central traits yielded the following results.

I. Interpersonal Perception (Tables 3&4)

Millon believes that the Passive-Dependent individual is perceived by others as being "gently affable," i.e., having rather kind and humanitarian impulses towards others. These individuals minimize their accomplishments and give notoriety to their inferiorities. They assume personal blame for problems that they have brought to others. Their friends view them as generous, cooperative, and thoughtful and note their generally apologetic and obsequious nature. In order to secure consistent care and affection from others on whom they are so dependent, Millon believes that these individuals must be admiring, loving, and willing to give their "all" for the other. They have learned how to behave affectionately towards others, and appear to possess a real capacity to give affection, respect, tenderness, and consideration. All of these interpersonal behaviors are perceived by others, and the Passive-Dependent person hopes that by behaving in such a manner that they will be protected, loved, and taken care of by an individual stronger and more able than themselves. Based upon the results from this study, only parts of Millon's hypotheses on the Interpersonal Perception of the Passive-Dependent personality could be supported. Rather than being outgoing towards others, they appeared to be introverted and anxious in social situations (16 PF, Factor A). They were not overly trusting (16 PF, Factor L). They did present themselves as meek, sensitive, and gentle (CPI, Femininity). They appeared to others as needing a greater amount of support and guidance (IBI succorance), and willing to take the blame when things went wrong (IBI abasement).

Millon believes that Compulsive individuals tend to restrain their affect, to be emotionally controlled, and perceived by others as grim and cheerless. They are portrayed as being rather austere and serious. Socially they appear rather stubborn, uncreative, and unimaginative. They seem most content with their nose to the grindstone, working on matters of detail and organization rather than interacting socially. These hypotheses were only partially supported in this research. The Compulsive sample was not socially austere and withdrawn (16 PF, Factor A), nor were they excessively rigid. (Rehfishch Rigidity Scales). They did appear to be

serious and ambitious (CPI Femininity), as well as presenting themselves as somewhat self-reliant (IBI succorance), not requiring the support of others.

II. Cognitive Functioning (Table 5)

Millon characterizes the cognitive functioning of the Passive-Dependent person as one of denial of these individuals' discomfoting thoughts about themselves and their strong dependency needs. These individuals constrict their world-view and limit their awareness of self and others to that which is within comfortable boundaries. There is an attempt to rationalize failues and the need for someone on whom to be dependent by blaming physical illness, unfortunate circumstances, or luck. The picture that emerged from the present study was that the cognitive denial of the Passive-Dependent person either fails easily or does not exist at all. They did not appear to be able to deny to themselves their feelings of anxiety and emotional lability (16 PF, Factor C). They appeared to proceed through life in a rather passive manner, seeing life in very somber terms (16 PF, Factor F). They seemed to believe that they are controlled by external forces (Rotter I/E) and do not have the ability to compete for themselves (IBI, Competition), and thus don't make the effort.

For Millon, the cognitive world of the Compulsive personality is one of constriction, of being rather narrow-minded and pendantic in thought. These cognitions are hypothesized to result from the contradictions in the personality organization of these individuals, wherein they would outwardly conform but internally be filled with unconscious impulses to act out in some socially disapproved manner. To protect himself, the Compulsive individual takes few risks and attempts to operate with some degree of certainty in his life so that his equilibrium is undisturbed. This constriction allows them to maintain a steadfast control over their emotional worlds. Millon's hypotheses on this central trait appeared to have gained some support from the present study. The Compulsive sample presented their cognitive world as one of emotional stability, maturity, and the ability to face reality in a calm and restrained manner (16 PF, Factor C). However, counter to Millon's concepts, the Compulsive sample did not appear to be sober, taciturn, and serious on Factor F of the 16PF. Both the Rotter I/E Scale and the Detachment and Competition scales of the IBI appear to reflect a great deal of control of self and a desire to claim responsibility for whatever positive reinforcements came their way rather than share credit with another.

III. Self-Image (Table 6&7)

Millon postulated that the passive-dependent person viewed himself as a rather weak, inferior and unworthy individual who is cooperative with others, is considerate of their feelings, and generally unambitious. The lack of dominance and command over others was demonstrated by the sample's low scores on the Dominance scale of the CPI, and was further confirmed by the raters scoring them low on the Dominance scale of the IBI. This would suggest a very apathetic, shy, unassuming, suggestable, and reserved set of individuals. This apparent lack of self-direction and control was further evidenced in their low scores on the Expressed Control dimension of the FIRO-B. While it was initially hypothesized that these individuals would manifest some degree of gregariousness, the data from the Expressed Inclusion dimension of the FIRO suggests that they are more reluctant to engage in an interpersonal liason.

The Compulsive person was believed to see himself as practical, prudent, moralistic, somewhat tense, and responsible for self. On Factor C of the 16PF the Compulsive sample presented themselves as being very conscientious, persistent, moralistic, and staid in their attitudes. In the present sample men scored significantly higher than women, regardless of personality style. The controlled, compulsive, and exactness of this samples' self-image was demonstrated well on Factor Q3 of the 16PF. These

individuals appeared to be ritualistic and demanding upon themselves and others. Both the CPI and IBI Dominance scales indicated that these individuals are able to control and direct others. This is confirmed by their scores on the Expressed Control dimension of the FIRO. Interpersonally they appeared to see themselves as expressing only a moderate interest in being included in social groups and functions as measured by the Expressed Inclusion scale of the FIRO. The raters' scores on the IBI Affection scale appeared to bear out their lack of affection and warmth towards others.

IV. Interpersonal Behavior (Tables 8&9)

According to Millon, the interpersonal behavior of the passive-dependent is characterized by compliance and a willingness to submit and oblige others, to be conciliatory and placating towards them in a sacrificing and clinging manner. Factor E of the 16PF did indicate a desire on the part of these individuals to be submissive, obedient, easily led, and accommodating to the wishes of others. However, they did not appear to be overly tender-minded (Factor I, 16PF), or sensitive. Neither the Wanted Inclusion or Affection scales of the FIRO were able to discriminate the two groups. However, both fell in the lower range of scores when compared to the normative population. Their interpersonal reliance was strongly evidenced on the Wanted Control scale of the FIRO, and it appeared that they wanted others to guide and control their actions and dominate them. The raters found that the passive-dependent person was prone to submit to the wants and demands of others, as their very high score on the Submissive scale of the IBI indicated. However, they would not generally defer to the advice and counsel of others (IBI Deference), but seemed to want to go it alone in pursuit of their goals.

Millon believes that the interpersonal world of the Compulsive individual is one of respectfulness and ingratiation towards superiors, while being formal and legalistic with subordinates. There was an attitude of assertion and dominance presented by this sample on Factor E of the 16PF. The Wanted and Affection scales from the FIRO did not discriminate the two personality styles, but both were low when compared to normals. The Wanted Control score from the FIRO strongly suggests that this personality style was most comfortable when others leave them alone and allow them to be self-directed. The IBI Deference scale scores for this group indicated that these individuals tended to defer to the person in a superior position. However, they do not really submit, as the IBI Submissiveness scale demonstrated. Rather, they defer and appear to go along, and then do what they feel is right in order to accomplish their superior's wishes.

SUMMARY

The results of this study indicated that many of Millon's initial hypotheses about these two personality styles were confirmed. However, some significant differences emerged. The passive-dependent individual was found to be interpersonally isolated and withdrawn, apparently fearful of further censure and rejection. They respond to events requiring a degree of self-reliance with emotional lability and distress, rather than competing or making attempts at mastery. The Compulsive individual did not appear to be dogmatic or rigid, in distinct contrast to the prior literature about their functioning. However, they did appear self-reliant and independent in their dealings with others, based primarily on their internal control orientation. However, they can be outgoing and comfortable with a small group of associates who are known and trusted. The results of this study add to the increasing amount of empirically derived evidence of various personality styles.

REFERENCES

- Cattell, R. B. The sixteen personality factor questionnaire (16PF). Champaign, Institute for Personality and Ability Testing, 1968.
- Diagnostic and statistical manual of mental disorders (3rd ed.). Washington, D.C.: American Psychiatric Association, 1979.
- Gough, H. G. California psychological inventory. Palo Alto: Consulting Psychologists Press, 1956.
- Lorr, M., Bishop, P. F., & McNair, D. M. Interpersonal types among psychiatric patients. Journal of Abnormal Psychology, 1965, 70, 468-472.
- Millon, T. Modern psychopathology: A biosocial approach to maladaptive learning and functioning. Philadelphia: Saunders, 1969.
- Olson, C. L. On choosing a test statistic in MANOVA. Psychological Bulletin, 1976, 83, 579-589.
- Rehfishch, J. M. A scale for personality rigidity. Journal of Consulting Psychology, 1958, 22, 10-15.
- Rotter, J. B. Generalized expectancies for internal versus external control of reinforcement. Psychological Monographs, 1966, 80 (Whole no. 609).
- Schutz, W. C. The interpersonal underworld. Palo Alto: Science & Behavior Books, 1966.

TABLE 1

Interpersonal Perception

Style C

"Gentle Affability--kind, soft humanistic, inadequate, responsible to others, obliging, grateful."

Style G

"Restrained Affectivity--emotionally controlled, grim, cheerless, respectful, duty-bound, moralistic."

16PF

(Cooperative, trusting, outgoing) Factor A (Reserved, critical, distrustful)
(Trusting, pliant, accepting) Factor L (Dogmatic, irritable, suspicious)

CPI

Feminity Scale (Fe)

Appreciative, helpful, gentle, respecting & accepting of others

vs

Hard-headed, ambitious as being manipulative and opportunistic in dealing with others

IBI

Successance Scale

Wanting, needing and expecting help from others

Perceived as being able to take care of self without external assistance

Abasement Scale

Willing to take the blame and responsibility when things go wrong

Perceived as being responsible but not as being the sole heir to mistakes or misfortunes

Rehfish R Rigidity Scale

Dependent, anxious, guilt-ridden, depressed

Obsessive-thinking, suspicious

Cognitive Functioning

"Cognitive Denial--refusing to admit discomforting thoughts about others"

"Cognitive Constriction--narrow-minded, overly methodical and pendant in thought"

Style C

Style G

16PF

(Easily upset, worrying)
(Enthusiastic, heedless, happy-go-lucky)

Factor C (Stable, mature, restrained)
Factor F (Sober, taciturn, serious)

Rotter I/E Scale

External Orientation

Internal Orientation

IBI

Detachment Scale

Perceived as being more
attached to groups and
persons

Perceived as being more de-
tached from groups and
persons

Competition Scale

Perceived as being less
competitive

Perceived as being more com-
petitive

Self-Image

"Inadequate--views self as
inferior, fragile, unworthy,
cooperative, considerate,
unambitious, pitiful."

"Conscientious--views self as
practical, prudent, moralistic,
responsible, tense, perfection-
istic."

Style C

Style G

16PF

(Disregards rules, expedient)

Factor G

(Conscientious, persistent,
moralistic, staid)

Factor Q3

(Uncontrolled, lax, follows
own urges, careless of social
rules)

(Controlled, exacting will-
power, socially precise,
compulsive)

CPI

Dominance Scale (Do)

Retiring, unassuming, inhibited,
avoiding situations of tension &
decision, lacking in self-
confidence, submissive, sug-
gestible, gentle

vs

Planful, confident, persistent,
persuasive, self-reliant, stern,
responsible, self-confident,
dominant, resourceful, ambi-
tious, forceful, planful,
stable

IBI

Dominance Scale

Perceived as unable to domin-
ate or influence others

Perceived as attempting to domi-
nate and influence others

Affection Scale

Perceived as being able to
give and receive affection
easily

Perceived as being unable to
give and receive affection
easily

Interpersonal Behaviors

"Interpersonal Compliance--
willing to submit & oblige
others, conciliatory, suggest-
able, placating, helpless,
sacrificing, clinging."

"Interpersonal Respectfulness--
ingratiating with superiors,
formal and legalistic with
subordinates, conforming,
polite, deferential, auto-
cratic, ritualistic."

Style C

Style G

16PF

(Submissive, dependent, humble)

Factor E

(Competitive, stubborn, aggres-
sive)

(Clinging, seeking help &
sympathy)

Factor I

(Taking responsibility, self-
reliant)

FIRO-B

High Wanted Inclusion Score
High Wanted Control Score
High Wanted Affection Score

IBI

Low Wanted Inclusion Score
Low Wanted Control Score
Low Wanted Affection Score

Deference Scale

Perceived as being less willing
to defer to the opinions of
those in authority

Perceived as being more willing
to defer to the opinions of
those in authority

Submissive Scale

Perceived as more willing to
submit to the wants of others

Perceived as less willing to
submit to the wants of others

TABLE 2

PILLAI'S TRACE CRITERIA FOR MANOVA HYPOTHESES

	<u>V</u>	<u>df</u>	<u>F</u>
Personality Style	0.9014	1	22.53***
Sex	0.6241	1	4.09***
Personality Style x Sex	0.2862	1	0.99

***p < .001

TABLE 3

ANOVA FOR VARIABLES OF INTERPERSONAL PERCEPTION

Scale		Personality Style				MS	df	F
		Style C		Style G				
		M	SD	M	SD			
Factor A	(16PF)	3.96	1.99	5.36	1.97	49.00	1	12.34***
Factor L	(16PF)	5.42	2.52	4.74	2.03	11.56	1	2.22
Femininity	(CPI)	22.80	5.20	18.50	4.65	462.25	1	24.59***
Succorance	(IBI)	33.56	5.08	16.24	5.42	7499.56	1	273.36***
Abasement	(IBI)	25.26	2.88	15.84	5.84	2218.41	1	107.04***
Anx in SS	(RRS)	9.44	3.21	4.80	3.41	538.24	1	44.78***
Need Struc	(RRS)	4.04	1.39	3.02	1.60	26.01	1	11.98***
Cons & Conv	(RRS)	5.14	1.57	3.90	1.57	38.44	1	15.57***

***p < .001

TABLE 4

ANOVA FOR SEX OF INTERPERSONAL PERCEPTION

Scale	Sex				MS	df	F
	Males		Females				
	M	SD	M	SD			
Femininity (CPI)	18.26	1.53	23.04	1.61	571.21	1	30.39***
Need Struc (RRS)	3.84	4.22	3.22	5.37	9.61	1	4.43*

***p < .001

*p < .05

TABLE 5

ANOVA FOR VARIABLES OF COGNITIVE FUNCTIONING

Scale		Personality Style				MS	df	F
		Style C		Style G				
		M	SD	M	SD			
Factor C	(16PF)	2.98	2.23	7.04	2.31	412.09	1	91.58***
Factor F	(16PF)	3.34	1.89	4.86	2.00	57.76	1	14.76***
Rotter I/E		12.24	4.27	6.44	3.75	841.00	1	46.97***
Detachment	(IBI)	23.08	6.74	26.82	6.90	349.69	1	7.57**
Competition	(IBI)	11.92	2.82	22.42	4.93	2756.25	1	168.18***

***p < .001

**p < .01

TABLE 6
ANOVA FOR VARIABLES OF SELF-IMAGE

Scale		Personality Style						
		Style C		Style G		MS	df	F
		M	SD	M	SD			
Factor G	(16PF)	5.20	2.10	7.98	1.45	193.21	1	65.13***
Factor Q3	(16PF)	4.10	1.97	8.70	1.23	529.00	1	181.11***
Dominance	(CPI)	15.50	5.77	30.12	6.60	5343.61	1	134.10***
Dominance	(IBI)	10.78	2.00	20.70	4.41	2460.16	1	202.69***
Affection	(IBI)	17.18	4.17	13.64	3.85	313.47	1	18.86***
Exp Incl	(FIRO)	2.34	2.46	3.96	2.47	65.61	1	10.57***
Exp Control	(FIRO)	0.82	1.31	2.94	2.75	112.36	1	23.62***
Exp Affect	(FIRO)	2.58	2.06	3.22	2.39	10.24	1	1.97

***p < .001

TABLE 7
ANOVA FOR SEX OF SELF-IMAGE

Scale	Sex						F
	Males		Females		MS	df	
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>			
Factor G (16PF)	6.94	1.95	6.24	2.53	12.25	1	4.13*

*p < .05

TABLE 8
ANOVA FOR VARIABLES OF INTERPERSONAL BEHAVIOR

Scale		Personality Style						
		Style C		Style G		MS	df	F
		M	SD	M	SD			
Factor E	(16PF)	3.48	1.97	6.34	1.86	204.49	1	55.24***
Factor I	(16PF)	6.08	2.19	5.68	1.86	4.00	1	0.95
Want Inc	(FIRO)	2.72	3.22	2.74	3.22	0.01	1	0.00
Want Control	(FIRO)	5.62	3.17	3.00	2.28	171.61	1	23.06***
Want Affect	(FIRO)	3.84	2.50	3.84	2.35	0.00	1	0.00
Deference	(IBI)	27.54	3.34	30.64	5.97	240.25	1	11.18***
Submissive	(IBI)	33.72	4.04	17.92	6.14	6241.00	1	195.01***

***p < .001

TABLE 9

ANOVA FOR SEX OF INTERPERSONAL BEHAVIOR

Scale	Sex				MS	df	F
	Males		Females				
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>			
Deference (IBI)	30.42	4.60	27.76	4.99	176.89	1	8.23*

*p < .05

THE RATIONAL - EMOTIVE DISPUTING TECHNIQUE:
AN APPROACH TO REDUCE PREJUDICE

BY

Z. BENJAMIN BLANDING, DEFENSE EQUAL OPPORTUNITY
MANAGEMENT INSTITUTE

Presented at the Eighth Psychology in the DoD Symposium, USAF Academy,
Colorado, April 22, 1982

Introduction

Civilized men/women have gained notable mastery over energy, matter, and inanimate nature generally, and are rapidly learning to control physical suffering and premature death. But, by contrast, we appear to be living in the stone age so far as our handling of human relationships are concerned (Allport, 1958).

In his book The Nature of Prejudice (1958), Gordon W. Allport explained that the world as a whole suffers from panic induced by that rival ideologies of east and west, each corner of the earth has its own special burdens of animosity. Today we still remain very conscious of moslems distrust of non-moslems. The children of Jews whose parents escaped extermination in central Europe still find themselves in the state of Israel surrounded by anti-semitism. As expressed by Allport some twenty-four years ago, refugees continue to roam in inhospitable lands and people of color suffer indignities at the hands of whites who invent and support racist doctrine to justify their condescension. "The checkerboard of prejudice in the United States is perhaps the most intricate of all," stated Allport. This checkerboard prevails today within our highly advanced, complex and scientifically intelligent society. Some of the barriers that block the acceptance of individuals are realistically founded within the bounds of conflicts of interest. However, I suspect that the most formidable foe that plagues the enhancement of our society in the areas of interacial/intercultural acceptance is prejudice.

When we speak of prejudice we are likely to think of "race prejudice." This is an unfortunate association of ideas, for throughout history human prejudice has had little to do with race. The conception of race is recent, scarcely a century old. For the most part, prejudice and persecution have rested on other grounds, often on religion. Until the recent past Jews have been persecuted chiefly for their religion, not for their race. Blacks were enslaved primarily because they were economic assets, but the rationale took a racial form (Allport, 1958). Today prejudice by its mere biased-based nature alone has become the roots of the racial formation.

What is Prejudice?

The question is asked, "what then is prejudice?" Gordon Allport investigated this word in great detail. He states that the word prejudice is derived from the Latin noun praejudicium, which has, like many words, undergone a change over time. The stages of transformation as reported by Allport are as follows: (New English Dictionary, Murray Ed. 1909):

- (1) To the ancients praejudicium meant a precedent - a judgement based on previous decisions and experiences.
- (2) Later, the term, in English, acquired the meaning of a judgement formed before due examination and consideration of the facts - a premature or hasty judgement.
- (3) Finally, the term acquired also its present emotional flavor of favorableness or unfavorableness that accompanies such a prior or unsupported judgement.

Dr. Allport concedes, when interpreting the definition of prejudice from the New English Dictionary that the wording recognizes positive as well as negative prejudice: A feeling, favorable or unfavorable, toward a person or thing, prior to, or not based on, actual experience. I agree with Dr. Allport that the biases of

prejudice may be pro as well as con. However, as Dr. Allport so clearly states, "It is nonetheless true that ethnic prejudice is mostly negative." The briefest definition presented by Allport is: Thinking ill of others without sufficient warrant. This definition is derived from the thomistic moralists who regard prejudice as "rash judgement." Allport gives credit for this definition to Rev. J.H. Fichter (LaFarge, 1945). Thus Gordon Allport in The Nature of Prejudice (1958), infers, implies, and hence indirectly explains that prejudice is an illogical, unempirical belief system causing irrational human thinking, emoting and behaving based upon irrationality. On the other hand, Albert Ellis in his book Humanistic Psychotherapy the Rational-Emotive Approach (1973), states that irrationality is any thought, emotion, or behavior that leads to self-defeating or self-destructive consequences which significantly interfere with rational thought processes, survival and happiness of an organism. In other words, giving due consideration to the principle of parsimony (Ockham's Razor), an illogical belief system is irrationality or irrationality is an illogical belief system. Illogical belief systems and irrationality are caused by overcategorizations; foolish, excessive demands; overblown generalizations and misconceptions made by individuals against other individuals or groups of people. Hence, prejudice can be and end product of human irrationality and an illogical belief system. Unfortunately, the span of this paper will not enable me to explore the social learning and biological basis of belief systems and human irrationality.

Allport (1958), even though not directly addressing belief systems and irrationality seems to be still more explicit about prejudice: The phrase "thinking ill of others" is obviously an elliptical expression that must be understood to include feelings of scorn or dislike, of fear and aversion, as well as various forms of antipathetic conduct: Such as talking against people, discriminating against them, attacking them with violence.

Similarly, we need to expand the phrase "without sufficient warrant." A judgement is unwarranted whenever it lacks basis in fact. A wit defined prejudice as "being down on something you're not up on."

It is not easy to say how much fact is required in order to justify judgement. A prejudiced person will almost certainly claim sufficient warrant for his views. He will tell of bitter experiences he has had with refugees, catholics, or orientals. But, in most cases, it is evident that his facts are scanty and strained. He resorts to a selective sorting of his own few memories, mixes them up with heresay, and overgeneralizes. No one can possibly know all refugees, catholics, or orientals. Hence, any negative judgement of these groups as a whole is strictly speaking, an instance of thinking ill without sufficient warrant.

Sometimes, the ill-thinker has no first-hand experience on which to base a judgement. A few years ago most Americans thought exceedingly ill of Turks - but very few had ever seen a Turk nor did they know any person who had seen one. Their warrant lay exclusively in what they had heard of the Armenian massacres and of the legendary crusades. On such evidence they presumed to condemn all members of a nation.

Ordinarily, prejudice manifests itself in dealing with individual members of rejected groups. But in avoiding a negro neighbor, or in answering "Mr. Greenberg's" application for room, we frame our action to accord with our categorical generalization of the group as a whole. We pay little attention to individual differences, and

overlook the important fact that negro X our neighbor, is not negro Y, whom we dislike for good and sufficient reason; that Mr. Greenberg, who may be a fine gentleman, is not Mr. Bloom, whom we have good reason to dislike. This process is so common that Allport might have defined prejudice as:

An avertive or hostile attitude toward a person who belongs to a group simply because he/she belongs to that group, and is therefore presumed to have the objectionable qualities as ascribed to the group (illogical, irrational belief system).

Finally, the point that Allport might have agreed with is that prejudice is indeed the result of unfounded, irrational beliefs and attitudes, lending itself to an already established, well proven therapeutic mode of treatment and becomes clear through his final definition:

Ethnic prejudice is an antipathy (irrationality) based upon a faulty and inflexible generalization (illogical and irrational belief system). It may be felt or expressed (irrational thinking and emotional consequences). It may be directed toward a group as a whole, or toward an individual because he/she is a member of that group (irrational behavior consequences).

RET and the RED Technique

Rational-Emotive Therapy (RET) shares the assumption, with other cognitive explanations of emotions and behavior, that the content of thoughts leads directly to emotional responses. In addition, RET makes another assumption: That rational beliefs typically lead to emotional and behavioral responses appropriate to personal goals, whereas irrational beliefs lead to sustained, dysfunctional emotional and behavioral responses. Consequently, one of the main goals of RET is to eliminate or greatly reduce irrational beliefs. RET is not a set of techniques; it is a theory of neurotic disturbance and a way of understanding people that may be applied to helping them change (Wessler & Wessler, 1980).

Since we have now seen that ethnic prejudice may be defined as irrational emotional and behavioral responses resulting from irrational thinking, it only appears rational that RET as an approach gives us a highly effective means to reduce prejudice. Rational-Emotive Theory (RET) makes effective use of a logic-based empirical process of scientific questioning, challenging, and debating the irrational beliefs and opinions that make up illogical, irrational belief systems. This scientific process may be referred to as making use of the Rational-Emotive Disputing Technique (REDT).

The REDT process is viewed as following some basic RET stages which Richard Wessler (1980), calls initiating RET, defining problems, and changing irrational beliefs. Much of the process is directed to the last stage, changing irrational beliefs. Initiating RET and defining problems may be accomplished rather rapidly. However, for the facilitator or counselor inexperienced in RET, defining problems ties in most directly with the theory of RET and may prove somewhat difficult at first (Wessler & Wessler, 1980). Certainly, training in RET theory and principles are necessary prerequisites.

First and most important, there is no one way to practice RET. Facilitators' styles in the use of Rational-Emotive Disputing Technique may range from a gentle,

evocative approach to a directly confronting approach. More often than not, a facilitator or counselor will use different styles, depending on the client or the group and what is happening during a particular individual session or group exercise. Thus REDT is not synonymous with a particular style, nor is it synonymous with particular techniques. REDT facilitators may differ in general techniques used, amount of activity, and degree of directiveness, although all to some degree would be both active and directive (Wessler & Wessler, 1980). Ruth and Richard Wessler maintain that directiveness, naturally, implies a direction. Directiveness is a very important factor when using this process to reduce prejudice.

A facilitator may conceivably use the same techniques as a RET therapist, yet not be practicing Rational-Emotive Therapy. Practicing RET means engaging in psychotherapy according to RET principles. Practicing REDT means using a set of techniques preferably in learning or a clarification group setting. REDT is to be used only as a tool to reduce or eliminate the kind of irrational thinking that produces problem emotion and behavior (prejudice, discrimination, hostility, anxiety, etc.) REDT is partly a systematic, specific, questioning technique directed not to confront the individual, but to confront an individual's manner of illogical thinking. Through the persistent application of questioning an individual's unfounded generalizations, a facilitator may guide an individual toward becoming aware of his/her inflexible, irrational line of thinking. With a renewed awareness of one's faulty manner of thinking, one would then be more susceptible to being open, more accepting of others as individuals, and less prejudiced.

REDT is a process of reeducation and relearning. More explicitly, REDT is used to undermine prejudice by adherence to the RET principles of disputing which involves debating, discriminating, and defining. If an individual maintains the irrational beliefs that all blacks are violent, dangerous, and oversexed, a facilitator may debate these irrational beliefs by asking rhetorical questions designed to rip up the false beliefs, such as: "What makes it so? What evidence supports it? In which way does it have truth - or falseness?" RET theory assumes, for example, that if an individual has a dysfunctional behavior or emotional consequence (discrimination and prejudice) following the occurrence of some unpleasant activating experience or event involving a black individual, that individual would almost certainly have some kind of belief that directly sparks this consequence, and that belief has irrational elements in it and therefore constitutes an irrational belief. The disputing method takes any shaky hypothesis, particularly one that leads to poor results, such as prejudice, and actively and vigorously disputes it, until it gets surrendered or sustained. If an individual finds evidence to back a hypothesis, fine; he or she keeps it. If one finds evidence against it, or if one does not find any to support it, one gives it up and looks for a better one. Debating, therefore, engenders an internal debate between one's rational and irrational beliefs. And the object of the debating? To destroy or at least minimize the irrational beliefs (Ellis and Grieger, 1977).

Discriminating consists of the facilitator assisting and teaching a student to distinguish between wants and needs, desires, and demands. Rational and irrational ideas. The process consists of showing an individual or a group the good as well as the bad points in behavior; noting the difference between undesirable and "unbearable" results in life; showing that hassles do not amount to "horrors"; differentiating between logical conclusions about life and non sequiturs; and discerning various other kinds of inconsistencies and contradictions

in thoughts and behaviors as one debates with oneself the irrational beliefs that produce prejudice (Ellis, 1977). Another aspect of disputing is the act of defining. Albert Ellis explains defining in RET as follows: (This definition is of course applicable to REDT). This scientific method usually starts off with some kind of definition of terms and ends up with even finer and more discriminative definitions. It tends to employ the principle of parsimony sometimes called Ockham's razor. Like the practice of general semantics, originated by Alfred Korzybski (1933), it particularly avoids and undercuts overgeneralizations, since they contradict some of the basic principles of logic.

Example of How It Can Work

Let's consider an example of how REDT can work. If you are prejudiced, you may feel an intense dislike towards a group or race of people. You resent them and your behavior towards them may range from avoidance to physical attacks. On one occasion a member of this group may have behaved badly towards you or a member of your family and thereby contributed to your prejudice. You tend to use several semantic absolutes of overgeneralization, when making reference to this group of people or one of its members and you probably tell yourself, irrational beliefs such as: (A) "All members of this particular group always act foolish, therefore I'll never learn to accept and understand them as equals", (B) "Now that I feel hostile towards this group of people, I can't stop feeling this way, and I'll continue feeling so forever", (C) "Because I am better than these people they should and must behave the way I'd like them to." The REDT practitioner would interrupt this kind of semantic looseness and induce you to debate, define, and discriminate your terms and statements more precisely, so that you see: (A) All members of a group of people do not always act foolish, and therefore you can, if you want to, learn to accept and understand them as equals, (B) You can stop feeling the hostility you now feel. You certainly need not feel that way forever, (C) Although it is good to have a high self-concept, it can be enhancing to recognize that the behavior of different individuals is usually the result of different sociocultural factors. It would be impossible for everyone to behave the same way. It is unnecessary for other groups of individuals to behave as you do in order for you to understand them. Think, how boring it would be if everyone behaved in the same manner.

REDT, in sum, consists of several kinds of cognitive restructuring, including focusing on irrationalities, debating against them, discriminating between logical thinking, and semantic defining and re-defining that helps one to stop overgeneralizing and stick closer to reality.

The Clarification Group Approach and REDT

The C-group was developed at the Boston University Center for Applied Social Science by Max Birnbaum in collaboration with Kenneth D. Benne, James Small, and various other colleagues. Its focus is on the dynamics of intergroup relations, and in this respect it differs from other small-group methodologies that focus on individual and inter-personal dynamics. Throughout the last twelve years, the C-group and a family of supporting technologies consistent with its learning objectives have been utilized in each of the center's annual laboratories in community relations and community dynamics, as well as in a variety of other settings in the United States.

The C-group is thus an experience-based learning methodology focusing on its

members' social origins as a means of understanding intergroup phenomena. It deals with stereotyping, prejudice, and distortions in perception, judgement, and behavior that are based on group identifications and intergroup conflict. C-group training is intended both to increase awareness of those processes and to train individuals to function more effectively with groups and individuals who are perceived as different from themselves.

The composition of a C-group learning community should be as heterogeneous as possible, unlike other small-group methodologies for which homogeneity of participants (at least along some crucial dimensions such as status, job, role, psychological sophistication, etc.) may be facilitative. If a group is to deal with a specific dimension, e.g., interreligious conflict, it is important that the composition of the group be representative of the religious groups that are in conflict.

The trainer in the C-group approach takes a directive role in leading the discussions and in the processing of experiences. This trainer behavior consists with the more traditional nondirective posture in most T-groups. Participants' previous group experience sometimes make it hard for them to adjust to and internalize the C-group style and to sort out what is relevant and what is irrelevant to C-group learning. It is the trainer's responsibility to direct the participants and enable them to function productively in the C-group. C-group training is, of course, not a totally structured experience in which free flow of expression is prohibited. However, the C-group does not feature the authority vacuum and the relevance of any emergent direction that characterize many T-groups and encounter groups. A by-product of this principle is the fact that the level of anxiety in the initial stages of the C-group is rather low compared with T-groups and encounter groups, and members experience anxiety differentially at later stages, depending on how they see the significance and emotional loading of specific contents. (Babad, Birnbaum, and Benne, 1978).

REDT may be effectively employed as a means to counteract prejudice, within the clarification group (C-group) approach. The general task of the REDT process in these groups would be what Wessler explains as dissuasive therapy, that is to dissuade people from their maladaptive ideas and unrealistic perceptions - in other words, to change attitudes.

Conceivably, REDT would greatly facilitate obtainment of the following C-group goals:

1. To reach awareness of one's own stereotypes, prejudices, and other distortions affecting one's social behavior. Via personal and interpersonal experiences in the C-group, one is also expected to become aware of the stereotypes, prejudices, and other distortions that are directed at one's own group by others.
2. To encourage realistic efforts to sort out valid and invalid elements in various stereotypes; to erase the dynamics of prejudice and to try to affect the more feasible causes of these; to understand and tone down the extreme polarizations that characterize intergroup conflict.
3. To analyze various instances of personal and interpersonal behavior and the reactions they elicit as representing stereotype behaviors of particular

groups and to deal with these behaviors as such. A higher order of this goal is to bring participants to a stage of analyzing a given event concurrently on two levels of analysis: A dynamic intra and interpersonal level and in intergroup level. It is important to understand that these levels of analysis are not contradictory, but complementary.

4. To gain sufficient mastery of theory and skills to become a more effective change agent in multigroup communities and other intergroup settings.

Summary

The Rational-Emotive Disputing Technique (REDT) method of science may be a very powerful, brief, interpersonal, institutionally employable approach to engender the surrender of those beliefs that ultimately cause or reinforce prejudice. Based on RET principles and theory, REDT may be successfully employed to counteract prejudice in a small group, nonclincical setting.

References

- Allport, G.W. The Nature of Prejudice. New York: Doubleday & Company, Inc. 1958.
- CF. A New English Dictionary. (Sir James A.H. Murray, ED.) Oxford: Clarendon Press, 1909, Vol. VII, Pt II, 1275.
- Babad, E.Y., Birnbaum M., and Benne, K.D. "The C-group Approach to Laboratory Learning." In J.E. Jone and J.W. Pfeiffer (Eds.) The International Journal for Group Facilitators, 1978, Vol. 3, No. 2, 168-184.
- Beck, A.T. Cognitive Therapy of the Emotional Disorders. New York: International Universities Press, 1976.
- Ellis, A. Humanistic Psychotherapy. New York: McGraw-Hill, 1974.
- Ellis, A. Reason and Emotion in Psychotherapy. Secaucus, N.J.: Citadel Press, 1977.
- Ellis, A., and Harper, R.A. A New Guide to Rational Living. Englewood Cliffs, N.J.: Prentice-Hall, 1975.
- Fabun, D. Communication: The Transfer of Meaning. Beverly Hills, CA. Glencoe Press, 1968.
- LaFarge, J. The Race Question and The Negro. New York: Longmans, Green 1945, 174.
- McGuire, W.J. "The Nature of Attitudes and Attitude Change." In G.L. Lindzey and E. Aronson (Eds.) The handbook of Social Psychology. Vol 2. Reading, Mass: Addison-Wesley, 1969.
- Meichenbaum, D. Cognitive - Behavior Modification: An Intergrative Approach. New York: Plenum, 1977.
- Wessler, R.L., "Evolution of Irrational Thinking." Rational Living, 1977, 12 (2), 25-30.
- Wessler, R.A. and Wessler, R.L., The Principles and Practice of Rational-Emotive Therapy. San Francisco, California: Jossey - Bass Inc., 1980.

MANAGEMENT ACTIONS TO REDUCE UA AND NON-EAS
ATTRITION IN THE U.S. MARINE CORPS

Ann Majchrzak, Ph.D.
Westat, Inc.

Abstract

Due to the significant negative effects on the U.S. Marine Corps of high rates of unauthorized absence (UA) and attrition, a three-year research effort was undertaken. This effort has two objectives: to identify specific unit-level practices for managing UA and non-EAS attrition; and to experimentally evaluate practices for reducing unit UA and attrition rates. The project, currently in its second year, has primarily focused on identifying unit-level management practices. In this endeavor, depth interviews coupled with a large-scale survey of commanders yielded numerous findings about management actions associated with lower unit UA rates. Stemming from these findings, experimental evaluation of specific management actions will be conducted in 1982.

MANAGEMENT ACTIONS TO REDUCE UA AND NON-EAS
ATTRITION IN THE U.S. MARINE CORPS

Ann Majchrzak, Ph.D.
Westat, Inc.

Statement of the Problem

High rates of unauthorized absence (UA) and attrition in the U.S. Marine Corps stand as barriers to readiness and to the timely deployment of manpower. Furthermore, significant incidence of violations impairs unit effectiveness and requires the investment of valuable leadership resources which further erodes a unit's ability to effectively fulfill its mission. Although there has been considerable research on identifying factors related to attrition and UA (e.g., Foch, 1979; Greenberg & McConeghy, 1977; Haber, 1975; Hoiberg & Berry, 1977), there has been little research focused on the management of these problems. Specifically, what has been lacking in the research efforts to date is the development of a set of activities which can be implemented at the unit level to reduce UA and attrition.

In response to this gap in previous research activities, a three-year research effort was undertaken in order to develop organizational interventions to reduce UA and attrition. This effort, which is currently in its second year, has two objectives:

- To identify specific unit-level practices for managing UA and non-EAS attrition; and
- to experimentally evaluate management practices for reducing unit UA and attrition rates.

To date, the project has been primarily focused on the first objective. Recommendations for experimental evaluations of management actions have been suggested based on the work-to-date. However, these recommendations are only now being implemented, with final experimental results expected to be available by early 1983.

In order to identify specific unit-level practices for managing UA and non-EAS attrition, two phases of data collection were undertaken. In the first phase (FY80), depth interviews were conducted with 32 problem and exemplary Marines and 32 members of chain of commands in eight units. The units included Division and FSSG battalions and Air Wing

squadrons on the east and west coast. In the second phase of the project, a survey of unit leaders was conducted to assess strategies for managing problem personnel. Battalion and company commanders in all 42 available Division and FSSG battalions in CONUS were surveyed in addition to commanders and staff NCOs in 35 randomly sampled Air Wing squadrons. The survey data were merged with data from headquarters on unit non-EAS attrition, UA, and desertions. The analysis-to-date has focused on identifying relationships between management factors and UA rates for unit leaders in the Division and FSSG units.

Findings of Depth Interviews

The depth interviews yielded a list of management factors that appear to influence individuals' decisions to go UA, or commit other violations. These management factors are:

- Lack of knowledge by unit commanders of the trends in UA and attrition for the unit;
- Inadequate integration of individual into unit;
- Inconsistent application of discipline by the unit chain of command;
- Limited knowledge of enlisted personnel about short- and long-term consequences of charge sheets;
- Limited options available to small-unit leaders for managing problem personnel;
- Inadequate knowledge of small-unit leaders about options available to help problem personnel;
- Problematic check-ins during entry into formal school and operational units; and
- Delays between check-in and training or assignment.

These factors must be considered tentative since they were not based on an empirical examination of a large sample of individuals.

Findings of Unit Leader Survey

The survey of unit leaders yielded several additional findings. The most important finding was a list of management factors which were found to be related to lower unit UA rates. Relationships were assessed by significant correlations between the factor and unit UA rate. The significant management factors are:

- Greater consistency among unit leaders in their actions and attitudes concerning the management of problem personnel.
- Greater knowledge among company commanders about unit trends in UA incidents.
- More information about off-duty activities provided to Marines by company commanders.
- Less endorsement of the expeditious discharge program as a useful management tool for handling problem Marines.
- More endorsement of informal punishments to handle first offenses.
- Less endorsement of involuntary discharges for Marines who repeatedly go UA.
- Less likely that "holding of charge sheets" will occur within the unit.
- More constructive management attitudes among company commanders, e.g.,
 - commanders perceive charge sheets as a last resort when a Marine violates a regulation;
 - commanders make time available to help Marines;
 - commanders perceive effective leaders as not resorting to discharges.

Furthermore, it was found that these management factors are highly interrelated. This high degree of interrelationship indicates that leaders of units with low UA rates exhibit more than one management action in order to keep the UA rates in their units low. Therefore, an optimal management strategy for reducing unit UA rates is one which incorporates several different management actions.

In addition to a list of management factors found to be related to lower unit UA rates, the analysis yielded several other findings. Some of those findings having impact on management of UA and attrition are listed below:

- Only seven percent of company commanders report UAs of less than one hour in length on the morning report. UAs between one and five hours in length are reported by only 38 percent of the commanders, while UAs of one day in length are reported by all company commanders. Therefore, in general, it seems that UAs must be a day in length before there is complete agreement on whether or not they should be reported.
- Company commanders felt that the most effective technique for handling first offenses (e.g., first UA, etc.) was the use of company-level NJPs. However, the analysis clearly indicated that the use of company-level NJPs for first offenses was not at all related to lower unit UA, desertion, or attrition rates. Rather, the techniques for handling first offenses found to be related to lower unit UA rates were the informal ones of extra training, informal counseling, and informal warnings.
- The average amount of days between the time a Marine is written up for a violation until some action is taken by the chain of command about the charge was reported by commanders as five days. This five-day period was surprisingly high since our earlier depth interviews had indicated that most commanders prefer that charge sheets be processed within a two-day period.
- Company and battalion commanders differ in the extent to which discharges are perceived to be a problem. Company commanders report that the number of early administrative discharges granted in their units represent a serious problem while battalion commanders report that the number of early discharges is only a minor problem.
- Both company and battalion commanders correctly expected that Marines who have committed one previous violation have a greater probability of committing another violation than Marines

who have not committed any violations. In fact, in analysis that we have performed on a 1978 cohort of entering Marines, we found that only 9.6 percent of the cohorts went UA for the first time during a one-year period while, of those Marines going UA once, 24 percent went UA a second time.

Recommendations

Based on the findings from the depth interviews and survey, we have made two major recommendations to the Marine Corps. Our first recommendation is that a management experiment be conducted to test the effects of a package of management actions intended to reduce unit UA and attrition rates. The actions to be experimentally evaluated may include such activities as accountability sessions with subordinate officers, the implementation of a management information system for commanders, and the development of group discussions with Marines on the application of discipline to specific problem cases. Our second recommendation is to disseminate to field commanders selected study findings which may be useful for managing Marines. Means of providing information to the commanders could include Leadership Training Courses, a Commandant's White Paper, or Leadership Field Pamphlets.

Preliminary discussions with Marine Corps Headquarters began in Fall, 1981. We plan to proceed with the first recommendation, with expected completion date of the experiment to be in early Winter, 1983.

References

- Foch, C. F. First-term enlisted attrition: Preliminary findings (RAND/WN-10267-MRAL). Santa Monica, California: 1979: The Rand Corporation.
- Greenberg, M.C., and McConeghy, G. "Exploratory development research of Navy and Marine Corps personnel" In W. Sinaiko, H.W. (Ed.). First-term enlisted attrition Proceedings of a conference held at Leesburg, VA, April 4-7, 1977. Volume I: Papers. Washington, D.C. Manpower Research and Advisory Services, Smithsonian Institution, 1977.
- Haber, S.E, Factors influencing attrition in the Marine Corps. (Serial T-306) Washington, D.C.: 1975, The George Washington University Institute for Management Science and Engineering.
- Hoiberg, A. & Berry, W.H. Fifteen years of research on the attitudes and performance of Marines. San Diego, California: 1977, Naval Health Research Center.

THE IMPACT OF MANAGEMENT VARIABLES ON ATTRITION IN UNITS

Mary Beth Merrin, Ph.D.
Westat, Inc.

Introduction

High rates of attrition among first-term enlisted soldiers have characterized the All-Volunteer Force. Although loss rates have decreased somewhat from a level of 40 percent among cohorts entering in FY74, they still remain substantial.

Research aimed at identifying factors which account for attrition has taken a variety of forms. Many efforts have focused on identifying individual background characteristics, such as age and education, which are related to attrition. (Mobley, Hand, Baker and Meglino, 1979; Manning and Ingraham, 1981; Berry, 1977). Others have documented the importance of biographical and personality variables as predictors of attrition (Wilcove, Thomas and Blankenship, 1979; Guinn, Kantor and Vithola, 1978).

A different line of attrition research has been developed which emphasizes the utility of a unit-level approach (Goodstadt and Glickman, 1974; Goodstadt, Yedlin and Romanczuk, 1978; Goodstadt and Romanczuk, 1980). This approach views attrition as a phenomenon which occurs at the unit-level and is influenced by factors in the unit environment. A recent study which adopted this approach found that unit-level factors (unit problems and strategies used by unit commanders in dealing with problem personnel) accounted for 30 percent of the variance in the administrative discharge rate of first-term enlistees, after controlling for the influence of education and unit type. (Goodstadt, Mallamad and Merrin, 1981.) The implications of these findings are dual. For one thing, they document the importance of unit-level factors in influencing attrition. In addition, these findings open up new areas for counter-attrition policies. On the basis of unit-level findings one can, for example, recommend changes which are implemented at the unit-level such as discharge and disciplinary regulations and practices employed by unit commanders.

A three-year research effort is currently underway which will make further use of the unit-level approach. This effort is now in the first year of implementation and is the subject of this paper. Because we are only in the data collection phase of the first year effort, this paper is necessarily in the form of an interim report on research methods and objectives.

Objectives of the Current Unit-Level Study

The major objective of our current research effort is to further identify management strategies used by members of the chain of command which are related to attrition among first-termers. We will also begin to assess the degree to which installation facilities can or do contribute to reduced attrition in units. These findings will be used to structure an experimental intervention aimed at testing counter-attrition strategies in the second year of the project. The results of the efforts undertaken in year one and two will then provide the basis for disseminating information in the form of a commanders' handbook on unit attrition management techniques.

Methodology of the First-Year Study

There are two major components to this year's research efforts:

1. Analyses will be undertaken using the OSD-ARI¹ unit level attrition data base. The focus will be on the impact of unit resources and personnel problems on loss rates at the unit and at the installation level.
2. Primary collection of qualitative and quantitative data will be undertaken to support analyses of the impact of unit resources and post facilities on unit and installation attrition rates.

Methodology for the Unit-Level Attrition Data Base Analysis

Analyses using the OSD-ARI unit-level attrition data base will be conducted in order to pursue several information objectives. One objective concerns the relationship between company commander attitudes, beliefs and problem-solving strategies to attrition within selected discharge chapters. The specific questions to be addressed here are:

¹The OSD-ARI unit-level attrition data base is composed of two major sub-files. One file contains information at the unit-level for the population of companies in CONUS. Included in this file is quarterly data on loss rates and educational composition. The other file contains aggregate information on losses at the company level combined with survey data from unit commanders drawn from 244 companies nested within 59 battalions on six installations.

- For each chapter category of losses, what is the relationship of company commander attitudes, beliefs and problem-solving strategies to attrition rates?
- Are certain categories of losses (e.g., Chapter 6) more predictable than others, given information on company commander attitudes, beliefs and problem-solving strategies?

A second information objective addresses attrition at the installation level and concerns differences in loss rates among installations. The specific questions to be addressed include:

- Are there differences in loss rates across installations?
- Are differences in loss rates across installations a function of problem incidence such that a higher incidence of problems is accompanied by a higher number of losses?¹

Methodology for the Primary Data Collection Effort

The primary data collection effort involves both qualitative and quantitative data collection. The qualitative data collection effort has been completed and involved conducting semi-structured interviews with battalion commanders, company commanders, company first sergeants, enlisted personnel and facilities administrators (ACS, AER, Red Cross, JAG, Morale Support, Mental Health, and Drug and Alcohol Abuse) at three installations. The objective of this effort was: 1) to help structure the quantitative data collection instrument; and 2) to provide a contextual base for later statistical analyses.

The quantitative data collection effort (currently underway) involves administering survey questionnaires at five installations. At each installation, questionnaire data is collected from the following respondents: 8-12 battalion commanders, 40-60 company commanders; 40-60 company first sergeants; 200 first-term enlistees; 7 facilities administrators. The sampling approach for collecting data at each installation is hierarchical such that we have information from all company commanders and company first sergeants within each battalion.

¹Data from the Sample Survey of Military Personnel will provide information on problem incidence.

The quantitative survey data will be merged with DMDC loss data at the company level. The subsequent analysis will address the following issues:

- 1) How does personnel management decision-making by the chain of command influence unit attrition (i.e., how much of the variance is accounted for by such variables, which managerial approaches are most closely related to attrition)?
- 2) To what extent do members of the unit chain of command make use of post facilities and how is such utilization related to unit attrition?
- 3) To what extent are first-termers aware of post facilities and to what extent do they utilize their services?

An Empirical Investigation of First Term Attrition

Glenda Y. Nogami

U.S. Army Research Institute for the Behavioral & Social Sciences
Alexandria, Virginia 22333

INTRODUCTION

First term enlisted attrition costs the Army hundreds of millions of dollars each year. In a 1980 Report to Congress, the General Accounting Office (GAO) determined that enlisted attrition from the 1974 to 1977 Cohort Years cost the government 5.2 billion dollars. These costs include the recruiting, training, and out-processing of early discharges, replacement recruiting, training and in-processing, in addition to veteran's and medical benefits to qualified early discharges (attritees). In addition to the high monetary costs, attrition also puts stress into the Army system. Personnel turbulence may affect unit readiness and unit effectiveness. Attrition creates a "domino effect" affecting recruiting (USAREC) and training (TRADOC) - i.e., more recruits and training for these new recruits is needed.

Although attrition is costly - as evidenced by both monetary and system stress, a certain amount of attrition is to be expected and welcome. No system of recruiting is perfect, so some "weeding out" of unqualified recruits is needed. What the optimum level of attrition should be has not been determined. What has been determined, however, is that the present attrition rate and attrition costs are too high.

When reviewing the literature on attrition, (c.f., Sinaiko, 1977; Goodstadt, Yedlin, 1979; Mobley, Hand, Baker and Meglino, 1979; Martin, 1977; Wiskoff, Atwater, Houle and Sinaiko, 1980; etc.) several things became clear. Differences in rates of attrition have been found between male and female soldiers (Ross & Nogami, 1981; Fox, 1979; Martin, 1979; Addington, 1979; and Tolk, 1978), between educational groups (Fox, 1979; Guthrie, Lakota, and Matlock, 1978; Manning & Ingraham, 1981), between female traditional and non-traditional MOS (Tolk, 1978; Ross & Nogami, 1981) and between stations of assignment (Whittenburg & Dahlinger, 1978).

Although much research has been done, there have not been any answers to questions such as: How does male attrition differ from female attrition? What is it about female traditional MOS which effects attrition? What are the differences between stations of assignment which affects attrition? How do the various demographic and biographic characteristics interact to affect attrition?

The views expressed in this paper are those of the authors and do not necessarily reflect the view of the U.S. Army Research Institute or the Department of the Army.

To systematically answer the above concerns, we at ARI have undertaken a four year programmatic research effort. We are now in the beginning of our second year. In this presentation, I would like to describe preliminary results from an on-going field research project.

Field Investigation of First Term Attrition

The objective of this research is to determine how attrition varies as a function of the characteristics of (a) the enlistee (i.e., demographics, reasons for enlisting, morale); (b) the MOS or job (i.e., traditionality of MOS, job environment, and competing civilian opportunities); and (c) location of assignment (i.e., Continental United States (CONUS)-Europe (USAREUR), well-liked vs. disliked post, availability of recreational and service facilities).

The resulting research design incorporates 2 geographic locations (CONUS and USAREUR); 2 levels of quality of installation (desirable vs. undesirable); 4 levels of MOS traditionality (female traditional, less traditional, non-traditional and combat) and gender (male and female). (See Figure 1).

		QUALITY	
		Desirable	Undesirable
	CONUS	**	**
	USAREUR	**	**

**contained within cell is the design below

		GENDER	
		Male	Female
TRADITIONALITY OF MOS	TRAD	CMF 31 71 (n=50)	CMF 31 71 (n=50)
	LESS TRAD	CMF 76 95 (n=50)	CMF 76 95 (n=50)
	NON-TRAD	CMF 63 64 (n=50)	CMF 63 64 (n=50)
	COMBAT	CMF 11 13 (n=100)	

Fifty first term enlisted males and fifty first term enlisted females in the traditional, less traditional and non-traditional MOS and one hundred first term enlisted males in the combat MOS were surveyed. In addition, one hundred first term enlisted early discharges and one hundred soldiers completing their term of enlistment were surveyed to provide comparison groups.

To provide a leadership perspective to first term attrition, 25 non-commissioned officers (E6 and above) and 25 company commanders were administered a leadership survey.

Results: Data collection in CONUS has been completed. Data collection in USAREUR is nearing completion. Consequently, the results presented will only cover the CONUS data, and these are preliminary findings.

Table 1 presents the total number of analyzable enlisted questionnaires in CONUS. The number of usable questionnaires from the early discharges and soldiers completing their enlistment was too low to allow a full factorial design. Until all data is in, analysis of these may be misleading. Consequently, the data presented will only be on first term enlisted soldiers still in the Army.

The preliminary results presented below were selected from over 130 questions. They were selected on two criteria: (1) there must be two or more results in the same direction and (2) acceptable significant levels must have been demonstrated. The results are presented in five sections: (1) reasons for enlisting, (2) work environment, (3) off-duty environment, (4) MOS (Job) characteristics, (5) availability of facilities.

		LOCATION			
		UNDESIRABLE		DESIRABLE	
		M	F	M	F
TRAD		45	34	41	48
LESS TRAD		44	47	35	34
NOT-TRAD		83	43	63	25
COMBAT		113		99	

TABLE 1

DEMOGRAPHICS: TRADITIONALITY OF CEF BY GENDER
BY DESIRABILITY OF LOCATION

1. Reasons for Enlisting: Career opportunities are more important to females than to males as a reason for enlisting.

1. Career opportunities in the military look better than those in civilian life. (19) (p=0.028)

Males = 2.38
Females = 2.24

2. I could make more money outside the Army. (70) (p=0.00)

Males = 4.03
Females = 3.62

3. A person can get more of an even break as a civilian than as a soldier. (51) (p=0.083)

Males = 3.52
Females = 3.36

2. Work Environment: A more desirable location is related to more satisfactory work climate.

1. All in all, I am satisfied with the soldiers in my work group. (58) (p=0.009)

Desirable = 3.34
Undesirable = 3.07

2. All in all, I am satisfied with the Army. (59) (p=0.02)

Desirable = 2.84
Undesirable = 2.64

3. All in all, I am satisfied with my unit. (60) (p=0.031)

Desirable = 2.55
Undesirable = 2.34

3. Off-Duty Environment: A. The installation we labeled as desirable is confirmed by respondents.

1. I want a reassignment to another post. (41) (p=0.073)

Desirable = 3.41
Undesirable = 3.73

2. All in all, this is a good post for me to live on. (61) (p=0.000)

Desirable = 2.69
Undesirable = 2.21

B. Females need more time off to take care of personal and family needs.

1. From the time you arrived at this installation, how many days have you been sick and could not work? (6A) (p=0.000)

Males = 2.12
Females = 2.86

2. I have enough time off to take care of my personal and family needs. (38) (p=0.035)

Males = 2.53
Females = 2.32

4. MOS (Job) Characteristics: A. Females have more of a mismatch between their PMOS and duty MOS than males.

1. MOS: I am working in my Primary MOS. (119) (p=0.000)

Males = 1.34
Females = 1.61

2. I am working in the job areas for which I have been trained. (121) (p=0.001)

Males = 1.26
Females = 1.39

3. What is your PMOS? What is your DMOS? (8,10)

Average Percent Mismatch

Males = 13.28%
Females = 27.65%
(see Chart)

B. Males spend more duty time in traditional male jobs (outdoors), females spend more duty time in traditional female (desk) jobs.

1. Outdoors (76) (p=0.000)

Males = 3.19
Females = 2.54

2. Doing heavy labor (77) (p=0.000)

Males = 2.45
Females = 1.86

3. Dangerous work (78) (p=0.000)

Males = 2.13
Females = 1.63

4. Dirty-Muddy-Oily work (79) (p=0.000)

Males = 2.90
Females = 2.20

5. Ash and Trash (80) (p=0.030)

Males = 2.27
Females = 1.96

6. Indoors (81) (p=0.000)

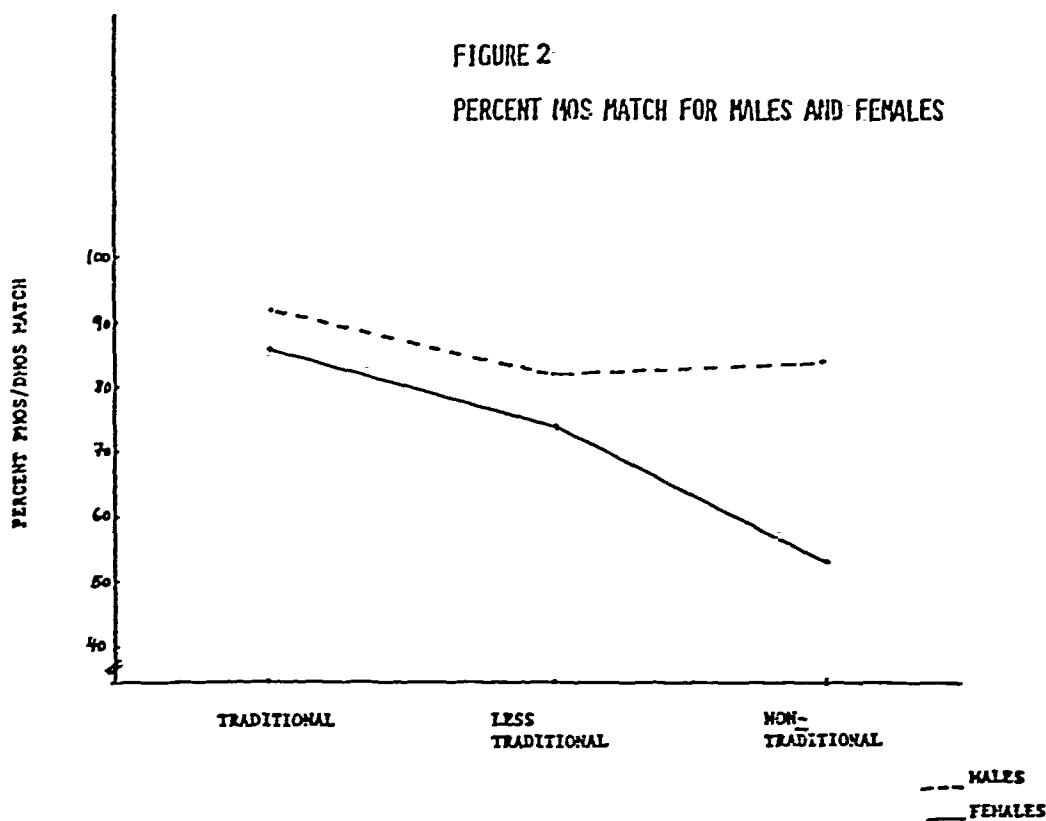
Males = 2.63
Females = 3.51

7. Doing paper work (82) ($p=0.000$)

Males = 2.48
Females = 3.45

8. Doing important work (83) ($p=0.001$)

Males = 3.15
Females = 3.61



5. Availability of Facilities: At both desirable and undesirable loc soldiers are satisfied with about the same total number of on-post pl off-post facilities.

Satisfaction with Facilities and Services		
Location		
	Undesirable	Desirable
On-Post	7.57	6.025
Off-Post	3.93	5.06

References

- Addington, J. Women's attrition: Establishing the relationship between high attrition and non-traditional skills. Information paper, DAPE-MPE-SS, 1979.
- Comptroller General (General Accounting Office). Attrition in the military- An issue needing management attention. Report to the Congress. FPCD-80-10, 1980.
- Fox, A. J. A comprehensive investigation of first term enlisted Army attrition. Draft report, Military Strength Programs Division, DCSPER, 1979.
- Goodstadt, B. E. and Yedlin, N. C. A review of the state-of-the-art research on military attrition: Implication for policy and for future research and development. Final technical report. Washington, D. C.: Advanced Research Resources Organization, 1979.
- Guthrie, N., Lakota, R., and Matlock, M. Voluntary release pilot program: Effects of attrition on general detail personnel. San Diego, California: NPRDC, 1978.
- Manning, F. and Ingraham, L. Personnel attrition in the U.S. Army in Europe. Armed Forces and Society, 1981, 2, 256-270.
- Martin, A. J. Trends in DOD First term attrition. In H. W. Sinaiko (Ed.), First term Enlisted Attrition - Vol. 1: Papers. ONR Technical Report 3, 1977.
- Mobley, W., Hand, H., Baker, R., and Meglino, B. Conceptual and empirical analysis of military recruit training attrition. Columbia, South Carolina: University of South Carolina Center for Management and Organizational Research, 1979.
- Ross, R. M. and Nogami, G. Y. The impact of MOS traditionality and soldier gender on first tour attrition. Draft ARI Technical report, 1981.
- Sinaiko, H. W. (Ed.) First-term enlisted attrition - Volume 1: Papers. Washington, D. C.: ONR Technical report, 1977.
- Tolk, A. B. The effects of MOS mismatch on females working in traditionally male and traditionally female MOSs. Final research report. Arlington, VA: Galler Associates, Inc., 1978.
- Whittenburg, J. A. and Dahlinger, N. E. Optimum tour length in USAREUR: First term enlisted personnel. ARI draft technical report, USAREUR, 1978.
- Wiskoff, M. F., Atwater, D. C., Houle, M. M., and Sinaiko, H. W. Enlisted first term attrition: Literature review and discussion. San Diego, California: NPRDC TR-80-, August, 1980.

ENLISTED ATTRITION: PROCESS AND OUTCOME

Introduction

First term enlisted attrition has been one of the problems that has surfaced in the All Volunteer Army. First term enlisted attrition costs the Army hundreds of millions of dollars each year. The costs incurred by attrition include the recruiting, training and out-processing of early discharges, replacement recruiting, training and in-processing, and veteran's and medical benefits to qualified early discharges. In addition to the high monetary costs, attrition puts stress into the Army personnel system, affecting rotation schedules, unit readiness and unit effectiveness.

Investigations of first term enlisted attrition include many different studies on a host of individual variables. For example, Goodstadt et al (1978, 1980) have considered management variables - usage of available command prerogatives; Fox (1979) and Sinaiko (1977, 1981) etc. considered the demographic characteristics of the enlistee's service experience, Tolk (1978) and Ross, Nogami, and Eaton (1982) have underscored the value of looking at the interaction of MOS traditionality and soldier gender.

The research reported in this paper tries to address this process from a person's decision to enlist to early discharge from the Army or until ETS.

METHODOLOGY

Research Design

Soldier Still in the Army. The research design incorporates 4 independent variables: geographic locations (CONUS and USAREUR), quality of installation (desirable vs. undesirable), MOS traditionality (traditional, less traditional, non-traditional, and combat) and gender. See Figure 2.

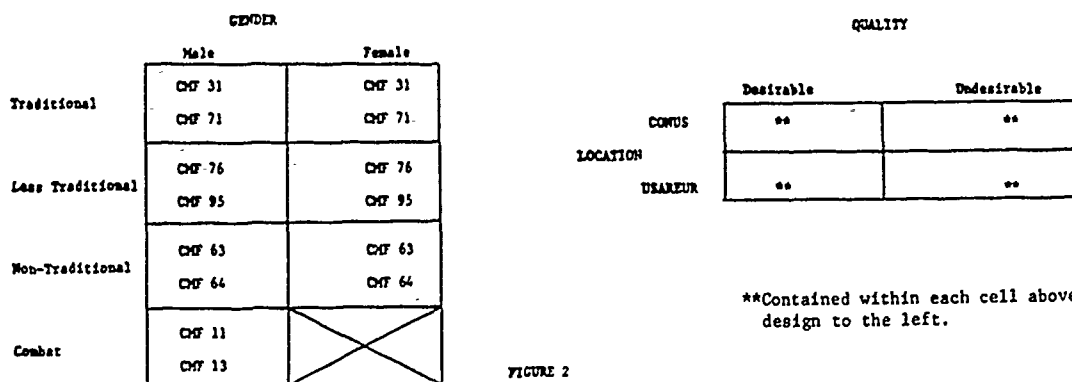


FIGURE 2

The data I will be presenting today will be only CONUS data - gathered from two separate FORSCOM installations.

Soldiers/Attritees/ETSers. In addition to the soldiers surveyed in the various traditionality categories shown in Figure 2, we attempted to survey attritees and ETSers on their way out of the Army. In this way, we could compare "successful" soldiers and "unsuccessful" soldiers. This resulted in a design with four independent variables: geographic locations (CONUS and USAREUR), quality of installation (desirable vs. undesirable), status of soldier (soldier, attritee, ETSer), and gender (see Figure 4).

		QUALITY	
		Desirable	Undesirable
COMUS	
LOCATION			
WEAREUR	

**Contained within each cell above is the design to the right.

FIGURE 4

	GENDER	
	Male	Female
SOLDIER		
ATTRITEE		
ETS		

Analysis. Responses to 62 multiple choice questionnaire items were obtained from all groups of soldiers. These 62 items were reduced to 18 factors through varimax rotation factor analysis techniques (see Figure 5). These factors were then analyzed through descriptive statistical (chi-square and frequency distribution) and analysis of variance techniques.

These 18 factors were:

Reasons for Enlistment

1. Personal growth
2. Military orientation
3. Job Opportunities
4. Distancing

Job Characteristics

5. Physical labor - like vs. dislike
6. Physical labor - percent of time spent
7. Unrelated job assignments - percent of time spent
8. Indoors, clerical work - percent of time spent
9. Undesirable work assignments

Attitudes

10. Unit identification
11. Army satisfaction
12. Civilian vs. Military options
13. Institutional orientation
14. Job advancement
15. Work load
16. Leadership concerns
17. Equal economic opportunity
18. Desirable posting

Figure 5

RESULTS

The results from this research will be presented in two sections. The first section will deal with problems or concerns that may surface through the process of the first enlistment. The sample for these analyses will be only those soldiers still in the Army. The analyses will be by traditional MOS, desirability of posting and gender. The third section will look at the concerns and problems as they may differentiate between ETSers, attritees, and soldiers still in the Army.

First Enlistment Process

1. Reasons for Enlistment

A. Personal Growth

Females were more likely to enlist for personal growth reasons than males. ($F = 7.646, 1/475 \text{ df.}, p = 0.006$)

B. Job Opportunities

Soldiers who are working in non-traditional female MOS felt that job opportunities the military offers were an important reason to enlist. This was least important for soldiers in the less traditional MOS. ($F = 2.984, 2/475 \text{ df.}, p = 0.052$)

2. Job Characteristics

A. Physical Tasks

-Overall, males spend more time doing physical tasks than females.

($F = 30.476, 1/495 \text{ df.}, p = 0.000$)

-Males also like spending more time on physical tasks than females.

($F = 8.348, 1/495 \text{ df.}, p = 0.004$)

-Soldiers in the non-traditional female MOS spend more time on physical tasks than either soldiers in the traditional or less traditional MOS. ($F = 26.090, 2/495 \text{ df.}, p = 0.000$)

-Soldiers in non-traditional MOS also like spending more time on physical tasks than soldiers in less traditional MOS who like physical tasks more than those in traditional MOS.
($F = 10.015$, 2/495 df., $p = 0.000$)

B. Unrelated Job Assignments

-Soldiers in the non-traditional MOS spend more time on unrelated work assignments than soldiers in traditional MOS.

($F = 2.911$, 2/495 df., $p = 0.055$)

-There are more perceived unrelated work assignments at the less desirable installation than at the desirable installation.

($F = 7.196$, 1/495 df., $p = 0.008$)

C. Indoor/Clerical Work

-Soldiers in the traditional and less traditional MOS spend more time indoors than soldiers in non-traditional MOS.

($F = 11.892$, 2/495 df., $p = 0.000$)

-Females spend more time indoors than males.

($F = 35.393$, 1/495 df., $p = 0.004$)

D. Job Match

-Males have a higher rate of PMOS and DMOS match than females.

-The less traditional the MOS category, the higher the PMOS-DMOS mismatch for females.

-Females are less likely to be working in their PMOS than males.

($F = 3.173$, 2/489 df., $p = 0.043$)

-Females are less likely to work in the job areas for which they have trained. ($F = 10.622$, 1/487 df., $p = 0.001$)

E. Equal Economic Opportunities

Females perceive more fairness in the system of job assignments than males. ($F = 22.024$, 1/493 df., $p = 0.000$)

3. Army Environment

A. Civilian vs. Military Options

Females perceive that the military offers more career/occupation advantages than the males. ($F = 12.730$, 1/482 df., $p = 0.000$)

B. Leadership Concerns

Soldiers on the desirable installation felt that their company commanders and NCO's cared about them more than soldiers on the undesirable installation. ($F = 15.281$, 1/484, $p = 0.000$)

C. On-duty vs. off-duty time

-Females took more sick days than males. ($F = 12.668$, 1/295 df., $p = 0.000$)

-Soldiers at the desirable posting took more sick days off than soldiers at the undesirable posting. ($F = 5.116$, 1/295 df., $p = 0.024$)

D. Desirable Posting

The desirable post was perceived as desirable.
($F = 25.035$, 1/493 df., $p = 0.000$)

STATUS DIFFERENCES

1. Reasons for Enlistment

There were no discernable differences between soldiers, ETSers, and Attritees, except as higher order interactions with desirability of posting and gender. Some of the n's in the higher order interaction cells were too small to base any conclusion on.

2. Job Characteristics

A. Physical Tasks

Attritees and ETSers like spending more time on physical tasks than soldiers. ($F = 9.834$, $2/915$ df., $p = 0.000$)

B. Unrelated Job Assignments

Attritees spent more time on unrelated job assignments than either ETSers or soldiers. ($F = 10.777$, $2/890$ df., $p = 0.000$)

C. Indoor/Clerical Work

Soldiers spent more time indoors than either ETSers or attritees. ($F = 3.048$, $2/915$ df., $p = 0.048$)

D. Job Advancement

ETSers perceived less promotion opportunities than attritees and soldiers. ($F = 5.519$, $2/896$ df., $p = 0.004$)

E. Job Match

-For males, there was a higher PMOS and DMOS match for ETSers than for attritees.

-For females, it appears that there is a better job match for ETSers than attritees. However the cells are very small (see Table 5).

-ETSers and soldiers were more likely to have a job match than attritees. ($F = 4.984$, $2/908$ df., $p = 0.007$)

-Overall, there was more of a job match at the desirable posting than at the undesirable posting.

($F = 5.505$, $1/908$ df., $p = 0.019$)

F. Equal Employment Opportunity

Attritees perceived that there was the least equal employment opportunities of all groups. In addition, ETSers perceived there was less EEO than the soldiers. ($F = 3.873$, $2/913$ df., $p = 0.000$)

3. Army Environment

A. Unit Identification

All three groups were significantly different from each other. The ETSers were the least satisfied with their units, the soldiers in the Army were the most satisfied. ($F = 10.860$, $2/890$ df., $p = 0.000$)

B. Army Satisfaction

All three groups were significantly different from each other. The attritees have the lowest Army satisfaction, soldiers had the highest satisfaction. ($F = 11.201$, $2/890$ df., $p = 0.000$)

C. Institutional Orientation

All three groups were significantly different from each other. The ETSers have the highest institutional orientation, followed by the soldiers, and then the attritees. ($F = 14.572$, $2/896$ df., $p = 0.000$)

D. Off-duty/On-Duty

ETSers felt they had enough off-duty time, attritees felt they had the least. ($F = 3.266$, $2/896$ df., $p = 0.039$)

E. Number of Dependents

a. Attritees have more dependents than ETSers and soldiers. ($F = 4.118$, $2/912$ df., $p = 0.017$)

b. Males are more likely to be single parents than females.

($F = 3.998$, $1/912$ df., $p = 0.046$)

Status The USAREUR data is currently being merged with the CONUS data. A final technical report is expected by late summer 1982. The final report will compare all those variables already mentioned above and include theatre differences and interactions among the variables.

REFERENCES

- Fox, MAJ. A. J. A comprehensive investigation of first term enlisted Army attrition. Draft report, Military Strength Programs Division, ODCSPER, Headquarters, DA, 1979.
- Goodstadt, B. E., Yedlin, N. C., and Romanczuk, A. P. Post-training enlisted attrition processes in the U.S. Army. Washington, DC: Advanced Research Resources Organization, 1978.
- Goodstadt, B. E., and Romanczuk, A. P. Research on determinants of unit level attrition decision making: I. Unit Commander Survey findings. OASD Technical Report, 1980.
- Ross, R. M., Nogami, G. Y., and Eaton, N. K. The impact of MOS traditionality and soldier gender on first tour attrition. Draft Technical Report, 1982.
- Sinaiko, H. W., Chatelier, P. R., Cook, C. A., Hosek, J. R., and Sicilia, G. T. (Eds.) Military personnel attrition and retention: Research in progress. Smithsonian Technical Report, 1981.
- Sinaiko, H. W. (Ed.) First term enlisted attrition- Vol. I: Papers. Washington, DC: Smithsonian Institution Technical Report, 1977.
- Tolk, A. B. The effects of MOS mismatch on females working in traditionally male and traditionally female MOSs. Final research report. Arlington, VA: Galler Associates, Inc., 1978.

A Preliminary Look at Vocational Interest
and Attrition Among Air Force Enlistees

Michael D. Matthews and Gene A. Berry

Air Force Human Resources Laboratory
Brooks Air Force Base, Texas 78235

Abstract

This study examines the relationship between predicted job satisfaction, as assessed by the Vocational Interest-Career Examination (VOICE), and turnover among first-term Air Force enlisted personnel. A large sample of recruits (N=49,668) was administered the VOICE during basic training and followed longitudinally through four years of active duty. A preliminary analysis of the data indicates a substantial relationship between predicted job satisfaction and attrition. The results suggest that a job assignment system which considers an objective assessment of predicted job satisfaction could reduce attrition from the Air Force.

Like most employers, the Air Force is concerned with obtaining the fullest possible utilization of its personnel resources. A critical part of that goal requires that incoming personnel be assigned jobs that will optimally utilize their abilities and interests. Present job-placement procedures rely primarily on the results of individual aptitude testing, job entry requirements, and needs of the service in evaluating suitability for competing assignments. An applicant's vocational preferences with respect to available jobs are typically assessed on a more informal basis during conversations with Air Force recruiting or occupational counseling personnel. Although some choice may be exercised on the part of the applicant during the process, decisions are sometimes made under less than optimal conditions. Since persons entering the service typically have little prior experience in the job market, and even less understanding of the Air Force occupational system, they understandably have a difficult time relating personal likes and dislikes to the choices available. The later consequences of misclassification at the entry level can be very costly for both the individual and the Air Force.

The Vocational Interest Career Examination (VOICE) is an Air Force instrument designed to assess vocational interests among Air Force enlistees. Its development and validation are described by Alley (1978). Job satisfaction, a key concept in models of employee turnover (Mobley, Griffeth, Hand, & Meglino, 1979), can also be predicted by the VOICE (Alley, Wilbourn, & Berberich, 1976). Job satisfaction has been found to be related to a variety of vocational behaviors including fatigue, dissatisfaction with life, depression, psychosomatic illness, mental illness, drug and alcohol abuse, job performance, and coronary heart disease (Alley, 1978). Perhaps the most serious implication of personnel dissatisfaction, however, has to do with its influence on various forms of occupational withdrawal. Research has demonstrated quite consistently that personnel dissatisfied with their jobs are much more likely to be absent from their work (Waters & Roach, 1973) and to terminate their employment at a higher frequency than are satisfied workers (Mobley et al., 1979).

The diverse and serious implications of job dissatisfaction led the Air Force Human Resources Laboratory to initiate a study of the relationship between vocational interests among first-term enlisted accessions, as assessed by the VOICE, and attrition from technical training and the Air Force. Also

being examined is the relationship between interests and retention, i.e., reenlistment for a second term of service. The purpose of the current paper is to describe preliminary findings summarizing the relationship between predicted job satisfaction, as assessed by the VOICE, and attrition from the Air Force. Data on other aspects of this study, such as the relationship between predicted job satisfaction and attrition from technical training, and reenlistment and extension rates, are currently being analyzed and are not yet available.

Method

Subjects

36,759 male and 12,909 female 1973-1975 Air Force enlisted accessions were administered the VOICE during basic training and tracked through their initial tour of duty. The subjects were typical of past Air Force accessions. Their average age was 18, the racial composition was similar to that of the population as a whole, and most (95.29%) had completed high school.

The VOICE

Form B of the VOICE was used. It consists of a 300-item vocational interest inventory requiring approximately 30 minutes to administer. Individual items are presented in booklet form and consist of occupational titles, work tasks, leisure time activities, and desired learning experiences. Respondents indicate relative preferences for each item in a standard like-indifferent-dislike (LID) format. Item responses were converted to two types of scales: (a) basic interest scales, and (b) occupational scales. The basic scales represent measures of general interest in various occupational and technical areas. They were constructed by grouping items of similar content into 18 independent sets covering a wide range of interests in the vocational and technical domain. The basic interest scales include measures of Office Administration, Electronics, Heavy Construction, Science, Outdoors, Medical Service, Aesthetics, Mechanics, Food Service, Law Enforcement, Audiographics, Mathematics, Agriculture, Teacher/Counseling, Marksman, Craftsman, Drafting, and Automated Data Processing. All items within each scale are homogeneous in the sense that each was selected to measure the same underlying dimension. The Office Administration items, for example, measure interest in clerical, administrative, and business related activities.

The occupational scales were designed for use in evaluating alternative areas of job assignment. Scores on the basic interest subscales were analyzed in depth and it was discovered that certain patterns of basic interest scores predict job satisfaction in various Air Force job clusters (Alley et al., 1976). These clusters, 20 in number, represent an exhaustive categorization of Air Force job specialties. The VOICE occupational scales provide a predicted job satisfaction score for each of these 20 job clusters. Consequently, job placement personnel are able to readily obtain a prediction of job satisfaction for any Air Force career field, by determining which of the clusters a particular job falls in. The scales, while based on interests, provide direct estimates of job satisfaction for each career field in the set and can be used for making specific comparisons between alternative assignments (Alley et al., 1976). Predicted job satisfaction (PJS) scores range from 200 to 800, with a mean of 500 and a standard deviation of 100. For a more thorough and technical discussion of the development of the VOICE and a description of the basic interest and occupational scales, their psychometric characteristics, and validity, see Alley (1978).

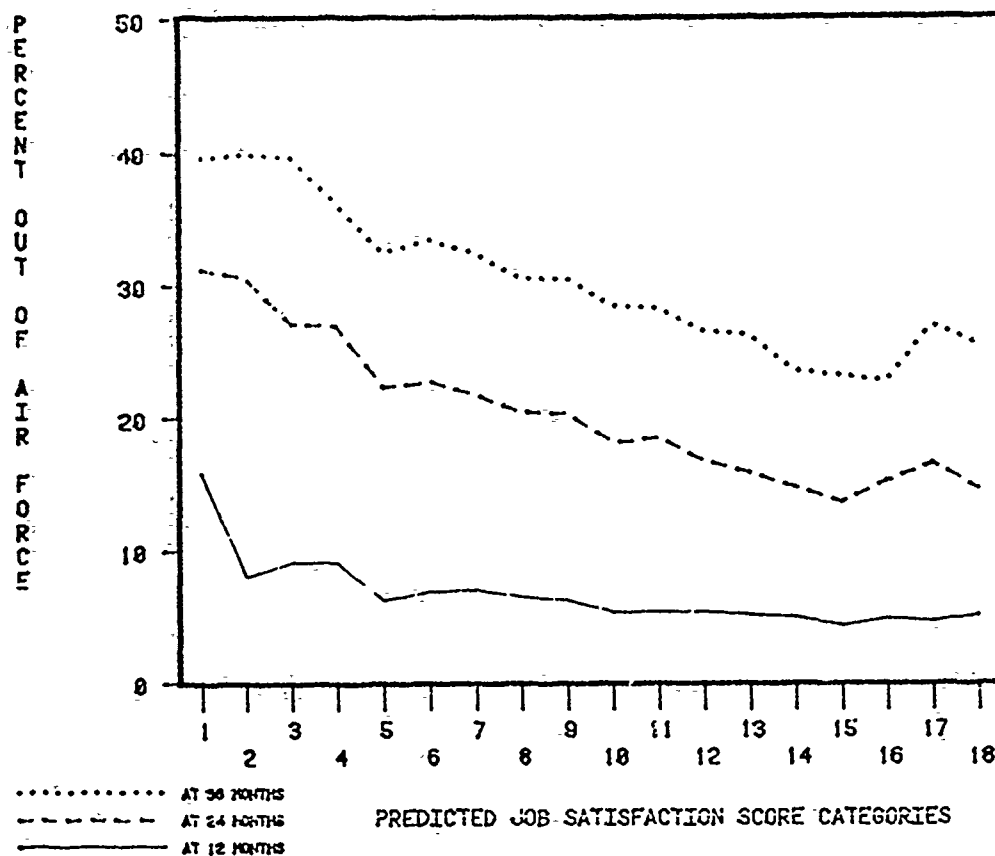


Figure 1. Percentage of subjects lost from the Air Force in 18 predicted job satisfaction categories (See Table 1 below) at 12, 24 and 36 months.

Table 1
Predicted Job Satisfaction (PJS) Categories

Category Number	PJS Range	N	Percent Out of Air Force		
			At 12 Mo.	At 24 Mo.	At 36 Mo.
1	Under 326	286	16.1	31.5	39.9
2	326 - 350	266	8.3	30.8	40.2
3	351 - 375	577	9.4	27.4	39.9
4	376 - 400	1017	9.4	27.2	36.4
5	401 - 425	1546	6.5	22.6	32.7
6	426 - 450	2089	7.2	23.0	33.7
7	451 - 475	2541	7.5	22.0	32.7
8	476 - 500	3171	6.8	20.7	30.8
9	501 - 525	4105	6.5	20.6	30.8
10	526 - 550	5370	5.6	18.4	28.7
11	551 - 575	6142	5.6	18.8	28.5
12	576 - 600	5950	5.6	17.0	26.8
13	601 - 625	5398	5.4	16.1	26.5
14	626 - 650	4376	5.2	15.0	23.8
15	651 - 675	3102	4.6	13.9	23.4
16	676 - 700	1884	5.1	15.5	23.0
17	701 - 725	713	4.9	16.8	27.2
18	Over 725	738	5.4	14.9	25.8

Procedure

The sample of recruits was allowed to complete their initial four to six year duty obligation. Cumulative attrition rates were assessed after 12, 24, and 36 months of service. Each subject's career field of assignment was identified and the PJS score associated with that field determined. Attrition rates were compared for each of the 20 VOICE DoD occupational clusters, and by sex within clusters. The occupational clusters were then combined for an overall analysis of attrition as a function of PJS score. Finally, these overall data were broken out by sex to examine possible effects of gender on the relationship between PJS scores and attrition.

Results

The relationship between predicted job satisfaction and attrition from the Air Force at 12, 24, and 36 months of service is depicted in Figure 1, which presents the percentage of cases lost from the Air Force within each of 18 groupings of PJS scores. For example, approximately 40 percent of subjects who had PJS scores under 326 (low predicted job satisfaction) had attrited within 36 months of their initial enlistment, versus 26 percent of the group with PJS scores over 726 (high predicted job satisfaction). Table 1 presents information on the PJS groupings. Preliminary tests of the overall relationship were performed using the Chi Square statistic. The results were Chi Square = 128.93, 287.40 and 230.58 for 12, 24 and 36 months respectively, with $df = 17$. These indicate a statistically significant ($p \leq .001$) relationship between predicted job satisfaction in the career field of assignment and attrition at all three time intervals.

A preliminary inspection of the data (not shown), broken down by sex, revealed no apparent difference in the relationship between PJS scores and attrition as a function of gender. Moreover, data within each of the 20 VOICE occupational clusters show no apparent systematic differences with respect to PJS scores and attrition, nor do results for the sexes appear to differ within clusters. However, more detailed analyses of the data are required to conclusively rule out differences among occupational clusters in the nature and magnitude of the relationships between PJS scores and attrition.

Discussion

While the relationships presented in Figure 1 appear fairly straight forward, several points should be kept in mind while interpreting the data. First, these data summarize attrition from all possible sources (e.g., voluntary, involuntary, death, hardship, pregnancy, misconduct, etc.). It is possible that predicted job satisfaction may be related to some, but not all, sources of attrition or that the magnitude of the relationship could vary among the various sources of attrition. Second, the distribution of PJS scores for the sample approximates a normal distribution with a mean of 560 and a standard deviation of 85. Consequently, the N's vary for each of the VOICE PJS groupings, ranging from smaller at the extremes to much larger in the center of the distribution.

This apparent relationship between predicted job satisfaction and attrition is consistent with results from an earlier study which found attrition from the Security Police and Law Enforcement technical school to be related to career interests as indexed by scores on the appropriate scale of the VOICE (Guinn, Wilbourn, & Kantor, 1977). These results suggest that Air Force enlisted personnel assigned to career fields in which they have high predicted

job satisfaction would have lower attrition rates than personnel assigned to career fields with low predicted job satisfaction. However, more rigorous statistical analyses of the data are necessary to determine the magnitude of the relationship between predicted job satisfaction and attrition, and to determine whether the use of predicted job satisfaction information would add to the predictive power of job assignment methods currently being used. Complex relationships between predicted job satisfaction and attrition which affect predictive power could exist. For example, it is possible that a low PJS score would reliably predict attrition above and beyond methods currently in use, while a moderate to high PJS score would not add to the power of currently used methods. If this were the case, some minimal PJS score should be viewed as a necessary, but not sufficient, condition for successful retention, and strategies for operational use of the VOICE would be affected.

The most important implication of these preliminary results would involve career field assignment strategies. Currently, vocational interests are considered informally in making career field placements. A recruit expresses a preference for one of a number of career fields in which aptitude scores qualify him or her for admittance and this information is considered along with the current manpower needs of the Air Force. The main problems with this system are that (1) the typical recruit, who has had little or no vocational experience, may not have a very well informed conception of how his interests relate to tasks performed in various Air Force jobs, and (2) abilities and manning needs overshadow the expressed interests of a recruit in assignment to a career field. An assignment procedure utilizing the VOICE could rectify some of these problems. It would permit consideration of vocational interest data along with manning needs and assessed abilities in making assignment determinations. In general, having looked at a large number of personnel assigned to their career fields without the benefit of predicted job satisfaction scores, it appears that substantial reduction in costly attrition could possibly be gained through consideration of such information.

REFERENCES

- Alley, W. E. Vocational Interest-Career Examination: Use and application in counseling and job placement. AFHRL-TR-78-62, Brooks AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory, October 1978.
- Alley, W. E., Wilbourn, J. M., & Berberich, G. L. Relationships between performance on the Vocational Interest-Career Examination and reported job satisfaction. AFHRL-TR-76-89, AD-A040 754. Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory, December 1976.
- Guinn, N., Wilbourn, J. M., & Kantor, J. E. Preliminary development and validation of a screening technique for entry into the security police career field. AFHRL-TR-77-38, AD-A043 919. Brooks AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory, July 1977.
- Mobley, W. H., Griffeth, R. W., Hand, H. H., & Meglino, B. M. Review and conceptual analysis of the employee turnover process. Psychological Bulletin, 1979, 86, 493-522.
- Waters, L. K., & Roach, D. Relationship between job attitudes on predictors of termination and absenteeism: Consistency over time and organizational units. Journal of Applied Psychology, 1973, 53(3), 341-342.

Personality Characteristics of Entering Air Force Academy Cadets

Richard L. Hughes

Abstract

The Personality Research Form - E was administered to entering male (N=520) and female (N=217) cadets in a recent class at the Air Force Academy. Group means for each personality trait were compared with civilian college norms for males and females. Results indicated that cadets as a group do represent a rather distinct personality profile compared to their civilian counterparts, and that male and female cadets are generally similar to each other in personality though they do exhibit distinct profiles when compared to their civilian normative groups. Comparison of male and female attritee's profiles with the whole class indicated that attritees do not reflect a unique group in terms of personality.

Many motives affect one's decision to enter a service academy and it is reasonable to think that personality is an important moderator affecting both the institution's appeal and an individual's successful competition for appointment. Besides having high academic standards, the academies emphasize athletics, training in military leadership positions, and motivational military experiences such as soaring or parachuting. Such aspects of academies, as well as long experience with cadets, suggested that on the average they may represent a distinct group personality profile. Furthermore, the admission of females to the academies in 1976 also raised the question of whether female cadets represented a similar personality profile as the males.

A related issue is who leaves the Academy? Do voluntary attritees share any personality traits which distinguish them from cadets who choose to stay? In the continuing attention paid to attrition/retention of cadets at USAFA, one school of thought has long held that the problem could be minimized by improving selection criteria. Stated differently, that position has viewed attritees as different from cadets who remain on some unidentified but critical variables. Therefore, in addition to examining the personality profiles of entering cadets, the present study also examined the profiles of resignees from the Academy to determine personality commonalities among them.

PROCEDURE

The Personality Research Form - E (Jackson, 1974) was administered to all females (N=217) and approximately one-half of the males (N=520) in an entering class at the Air Force Academy. They completed the test during the first few days of basic cadet training. Testers insured the respondents that the results would be treated anonymously and would not result in any actions toward them individually. Average profiles for incoming male and female cadets were computed from this group. Attritees from the class fourteen months later were also identified. That time represented the beginning of the class's sophomore year, and most attrition occurs by then. Attritees for academic or medical reasons were excluded, leaving a total of 453. It should be noted that a small portion

of this attrition was not truly voluntary (i.e. military aptitude dismissals, honor violations) but was included anyway because of the apparent relevance of these attrition categories to personality. After the academic and medical cases were eliminated, PRF scores were available for 106 of these (80 males and 26 females). The reduced number is because not all incoming cadets completed the PRF; it reflects a random sample of approximately 35% of the attritees.

RESULTS

Means and standard deviations for each of the PRF scales were computed for the male and female groups. Mean scores were compared with Jackson's published norms for college students. The results are summarized in Table 1.

Relative to civilian college females, entering female cadets were significantly higher in Achievement, Affiliation, Cognitive Structure, Dominance, Endurance, Exhibition, and Order. Female cadets were significantly lower in Harmavoidance, Play, Sentience and Succorance.

Relative to civilian college males, entering male cadets were significantly higher in Achievement, Affiliation, Cognitive Structure, Dominance, Exhibition, Nurturance, Order, Social Recognition and Succorance. Male cadets were significantly lower in Autonomy, Change, Harmavoidance, Sentience and Understanding.

A second analysis of the scores was based upon interpreting both male and female cadet PRF scale means against college male norms. We wondered what the new female cadets were like as a group, not only in terms of civilian female norms, but also, as it were, in terms of "cadet" norms. To answer that we looked at the placement of male and female cadet mean scale scores relative to the college male scale means. Male and female entering cadets were concurrently higher or lower relative to the college male means on 18 of 21 personality dimensions. They differed on the following scales: Aggression (females lower); Change (females higher); and Sentience (females higher).

Comparisons were also made for each trait between the incoming male and female group means and attritees in each group one year later. There were few differences. Male attritees were significantly higher in Aggression and lower in Abasement. Female attritees were not significantly different on any trait. In general, then, the personality profile for attritees is extremely similar to that for cadets who remain at the Academy.

DISCUSSION

The mean profile for entering cadets (male and female) seems easily interpretable in terms of an individual's attraction and selection to the Air Force Academy. It is an institution with an explicit achievement ethic and it attracts achievers. Its charter is to develop military leaders and it attracts relatively dominant or assertive individuals. Academy training emphasizes teamwork and the institution attracts gregarious personalities. Parts of training involve physically exciting and sometime frightening activities and entering cadets score low on Harmavoidance. A military Academy is almost a professional school at the undergraduate level and the individuals who elect the Academy experience are pragmatic and serious-minded. High in need for social recognition, cadets selected a school

and career where decorum and making a good impression are highly valued. Enjoying a structured environment and being systematic by nature, cadets attend an institution epitomizing consistency, rules, orderliness and planfulness.

It also seems clear that those who leave the Academy are very similar in personality to those who remain. In that sense, at least, it is difficult to see how a more refined selection procedure could reduce attrition. In terms of personality, the attritees did not seem to represent "selection errors."

In conclusion, two ambiguities of the study should be noted. One analysis involved comparisons between entering Academy cadets and college students from presumably a range of academic classes. It is thus possible that differences between the cadet means and civilian means reflect simply developmental differences. That is considered unlikely because the personality profile for entering cadets does seem conceptually so congruent with unique aspects of the institution. Another ambiguity concerns whether the circumstances of testing might have biased the results. Despite attempts to minimize that influence, it is nonetheless possible that the new cadets responded as they thought a cadet should respond. It is also possible that the stressfulness of the new environment, not the least of which involved recent separation from home for many, had an unknown effect on responses.

PRF SCALE MEANS FOR
AIR FORCE ACADEMY CADETS AND CIVILIAN COLLEGE STUDENTS

	Civilian College Females <u>N=115</u>	USAFA Incoming Females <u>N=217</u>	USAFA Female Attritees <u>N=26</u>	Civilian College Males <u>N=100</u>	USAFA Incoming Males <u>N=520</u>	USAFA Male Attritees <u>N=80</u>
Abasement	7.7	7.7	8.0	7.8	7.7	7.1
Achievement	10.0	12.4	12.4	11.0	12.2	11.9
Affiliation	8.9	9.9	11.1	8.3	10.5	10.6
Aggression	6.9	6.8	6.9	7.4	7.9	8.3
Autonomy	7.1	6.5	6.9	9.5	6.0	5.8
Change	9.9	10.1	10.2	9.5	8.6	8.8
Cognitive Structure	8.7	10.0	9.8	8.6	10.7	10.9
Defendence	6.0	5.6	6.1	5.8	6.0	6.4
Dominance	7.6	11.7	11.4	10.2	12.5	12.5
Endurance	10.2	11.2	11.2	10.0	11.5	11.2
Exhibition	7.2	8.3	9.3	7.5	8.5	8.8
Harmavoidance	9.5	6.0	5.4	7.4	5.6	5.7
Impulsivity	6.5	5.7	6.1	5.5	5.3	5.4
Nurturance	10.9	11.2	11.5	8.9	10.4	10.4
Order	8.2	9.3	9.4	7.8	8.9	9.0
Play	9.0	8.3	8.5	8.2	8.7	9.3
Sentience	10.8	9.9	10.4	9.3	8.4	8.4
Social Recognition	8.2	8.9	8.6	7.5	9.2	9.5
Succorance	8.7	7.0	7.7	5.6	7.3	7.5
Understanding	9.7	9.6	10.2	10.3	8.8	8.3

Table 1

REFERENCES

Jackson, D. N. Personality Research Form Manual. Goshen, New York: Research Psychologists Press, Inc., 1974.

Determining Officer Education Requirements

Lynn M. Scott
Air Force Human Resources Laboratory, Brooks AFB TX

Abstract

This study used the judgment analysis technique to examine the education requirements for seven Air Force officer specialties. Incumbents' holistic suitability judgments were gathered for 50 college transcripts presented in a standard format to each of 100 judges per specialty. Regression models based on four different predictor configurations were generated for each specialty and by incumbent rank within each specialty. Results showed that education suitabilities were constant across the officer ranks surveyed in all but two specialties and the optimal regression model to predict suitability used 12 predictors representing both linear and curvilinear effects.

The Air Force's use of officers is governed by a classification system which matches the officer's characteristics (education, training, and experience) with the job's requirements. Accurate matches are facilitated by an applicant pool with as much depth in these characteristics as possible. Depth in Air Force training and experience is accurately measured by documented training and assignment histories. However, the college transcript does not provide a similar metric of an officer's attained education. Establishing such a metric would provide the opportunity to quantify applicant pool characteristics, thus, improving classification actions.

Education requirements in the officer classification regulation (AFR 36-1) are usually expressed as an academic major or as a baccalaureate degree with emphasis in specific academic disciplines. A number of problems, however, surface when a personnel manager reviews a transcript for qualification. Goody (1977), determined that at least three limitations existed. They were: the absence of a valid method of distinguishing bachelor of arts degrees from bachelor of science degrees; the inability of academic major requirements to permit a range of reliable suitability assessments and; the inability to accurately equate the course content of transcripts from different colleges. His exploration of the problem continued with a pilot study to develop a format to standardize college transcript data and to evaluate the feasibility of using it in conjunction with judgment analysis (Christal, 1963) to predict education requirements. The results were successful and were followed by a larger study which generated education requirements for 10 specialties. Research to enhance these findings began in 1978. A major goal of this effort was the redesign and testing of a more useful method of standardizing transcript data. The document (Figure 1) used to code the data (called an Officer Education Profile) was revised to contain 48 academic course titles aggregated into six academic categories. A field test of the profile revealed that college transcript data could be coded into the education profile with minimal error. The purpose of this study was to survey education suitability ratings for 50 different education profiles from job incumbents in 12 specialties. Judgment analysis (Christal, 1963) was used to develop linear regression models of the incumbent's holistic judgments and; to assess each model's statistical and practical efficacy to represent education requirements.

Method

Specialties Surveyed

This study considered 12 Air Force Officer specialties (hereafter specialties) which were representative of the line officer occupational structure. Three specialties were selected from each of the following four categories: rated (flying) specialties that were specifically combat related (e.g., tactical fighter pilot); rated specialties that were not specifically combat related (e.g., airlift navigator); support (non-flying) specialties that were specifically combat related (e.g., air weapons controller) and; support specialties that were not specifically combat related (e.g., acquisition program manager). The judgment analysis technique (Christal, 1963) was used to determine which combination of semester hours would best predict the average suitability rating. A survey solicited specialty incumbents' relative suitability ratings for 50 college transcripts coded in a standardized education profile.

Subjects

The subjects were 1129 officers (First Lieutenant through Major) randomly selected from the Air Force Uniform Officer Record. Subjects represented the following 12 specialties: Tactical Fighter Pilot, Strategic Bomber Pilot, Missile Launch Officer, Transport Pilot, Airlift Navigator, Tanker Pilot, Air Intelligence Officer, Munitions Officer, Air Weapons Controller, Scientific Analyst, Administrative Management Officer, and Acquisition Project Officer. The approximate sample size per specialty was 100.

Education Profiles

Fifty baccalaureate level transcripts were selected which represented the proportion of academic majors produced in the 1976-77 academic year. Each transcript was coded into the education profile format and displayed in an abbreviated format which only listed rubrics with non-zero semester hours. The profiles were randomized in four unique orders to minimize stimulus context effects.

Apparatus

Each subject received a survey composed of the following materials: background questionnaire which solicited demographic and job properties data; a nine-point relative suitability scale; a blank copy of the officer education profile; definitions of the 48 academic titles listed in the profile; optically scannable response sheets; an education profile booklet of the 50 transcripts; a return envelope; a "time to complete" question and; instructions to complete the survey.

Procedure

Surveys were packaged for direct mailing to the subjects. The four uniquely ordered education profile packages were evenly distributed across subjects in each specialty. The subjects' principal task was to assign the relative suitability rating that best described their impression of usefulness for the education reflected in a profile. Each profile was to be

viewed as the educational background of a possible successor to them in their job. A three month time period was given for survey administration.

Data Analysis

Four multiple linear regression models were developed to test the independent and multiplicative effects of the 48 academic title semester hours and their aggregation into six academic categories on the criterion, the education suitability rating. The regressions were configured in the following manner for each specialty: Model 1, a full model containing all 48 academic title semester hours as predictors; Model 2, another full model containing 27 predictors consisting of the six academic category totals, the category totals squared and all possible category by category interactions; Model 3, a 12 predictor restricted model consisting of academic categories and their squares; Model 4, also a restricted model with the six academic categories as predictors. Figure 1 shows the predictors subsumed in each model. Each model was assessed by grade within each specialty in addition to its assessment for all respondents within each specialty. This design addressed two hypotheses. The first hypothesis was that within a specialty, there is no difference in regression coefficients across grades. The second hypothesis was that there is no difference in predictability between any of the full and restricted models.

Results

The overall survey return rate was 55%. Across specialties, rates ranged from 69.5% to 33%. Insufficient returns across grades in five specialties caused them to be eliminated from further study. The following seven specialties were retained for additional analyses: Tactical Fighter Pilot; Transport Pilot; Airlift Navigator; Tanker Pilot; Air Weapons Controller; Scientific Analyst and; Acquisition Project Officer.

Grade was a significant effect only for the Transport Pilot, Tanker Pilot, and Acquisition Project Officer specialties ($p \leq .01$). That is, in these three specialties, officers of different grade gave different suitability ratings for the same profile. The Bottenberg and Christal (1961) pooled R^2 value was computed to determine the overall efficiency with which the group scores would have been predicted if grade specific equations had been used. The pooled R^2 was compared with the decreased R^2 value for the non-grade specific regression and differences were tested with the F statistic defined in Gott (1978). Follow-on analyses addressing predictability across regression models did not include grade effects for any of the specialties.

The most efficient regression model for Air Force specialty was selected using the F statistic computed from R^2 values for full and restricted models described in Bottenberg and Ward (1963). Four full vs restricted model comparisons were made for each specialty: (1) the 48 predictor full model (Model 1) against the 6 predictor restricted model (Model 4); (2) the 12 predictor full model (Model 3) against the 6 predictor restricted model; (3) the 27 predictor full model (Model 2) against the 12 predictor full model; (4) the 27 predictor full model against the six predictor restricted model. Model 1, the 48 predictor model, was most efficient for the Transport Pilot, Airlift Navigator, Tanker Pilot, Scientific Analyst and Acquisition Project Manager specialties. Model 3, the 12 predictor model, was most efficient for the Tactical Fighter Pilot and Air Weapons Controller specialties.

Two approaches were taken to interpret education requirements from the regression statistics. The first approach compared the regression coefficients. The second approach examined the sequence that the predictors entered the regression equation. This sequencing is interpreted as the rank ordering of the predictors' contribution to the R^2 value. For illustration, the 12 predictor model (Model 3) regression weights and predictor sequencing will be used to describe the education requirement for the Acquisition Project Officer Specialty. This specialty's 12 predictor regression is $y = 1.17 + .1141 (\text{Total Business Hours}) - .0011 (\text{Total Business Hours})^2 + .0632 (\text{Total Math Hours}) - .0005 (\text{Total Math Hours})^2 + .0213 (\text{Total Social/Behavioral Science Hours}) - .0002 (\text{Total Social/Behavioral Science Hours})^2 + .0918 (\text{Total Engineering Hours}) - .0008 (\text{Total Engineering Hours})^2 + .0061 (\text{Total Physical Science Hours}) - .00009 (\text{Total Physical Science Hours})^2 - .0087 (\text{Total Humanities Hours}) - .0001 (\text{Total Humanities Hours})^2$. Its coefficient of determination is .49. The predictors that account for 99% of the explained variance are: (1) Total Business Hours; (2) Total Engineering Hours; (3) Total Math Hours; (4) $(\text{Total Engineering Hours})^2$; (5) $(\text{Total Business Hours})^2$; (6) Total Physical Science Hours. It can be interpreted that for this specialty, courses in business, math, engineering, and some physical sciences are the most essential; while noteworthy curvilinear effects exist for large numbers of business courses and, more importantly, engineering courses.

Discussion

Although statistically superior models were identified, practical differences across models were absent. That is, for each of the specialties, there was no practical difference in R^2 across the models. With this in mind, a single model with specialty specific regression weights can be used to derive education suitability. Model 3, the 12 predictor regression would optimally serve this purpose. This model's predictors are the semester hour totals for the six academic categories and the squares of each category total. The model permits a clearer interpretation of regression coefficients than the coefficients in the 48 predictor model. An additional advantage is its reduction of the number of cautions associated with overprediction when using models with a large number of predictors. Finally, Model 3 can account for curvilinear effects; a capability missing from the six predictor model tested.

In all, this study's findings have demonstrated that education suitability ratings can be gathered from incumbents and replicated by the judgement analysis technique.

References

- Bottenberg, R. A., & Christal, R. E. An iterative technique for clustering criteria which retains optimum predictive efficiency. WADD-TN-61-30, AD-261 615. Lackland AFB TX: Personnel Laboratory, Wright Air Development Division, March 1961. Also Journal of Experimental Education, 1968, 36(4), 28-34.
- Bottenberg, R. A., & Ward, J. H., Jr. Applied multiple linear regression. PRL-TDR-63-6, AD-413 128. Lackland AFB, TX: 6570th Personnel Research Laboratory, Aerospace Medical Division, March 1963.

Christal, R. E. JAN: A technique for analyzing group judgement.

PRL-TDR-63-3, AD-403 813. Lackland AFB, TX: 6570 Personnel Research Laboratory, Aerospace Medical Division, February 1963.

Goody, K. Matching job education requirements of a variety of officer

specialties with the educational attainments of potential incumbents.

AFHRL-TR-77-44, AD-A050-826. Brooks AFB, TX: Occupation and Manpower Research Division, Air Force Human Resources Laboratory, August 1977.

Gott, C.D. HIER-GRP: A computer program for the hierarchical grouping of

regression equations. AFHRL-TR-78-14, AD-AB58 415. Brooks AFB, TX:

Computational Sciences Division, June 1978.

A Hybrid Organizational Development Strategy
for Small Group Interventions

Robert C. Ginnett

United States Air Force Academy

The primary strategy for job redesign interventions within the Air Force was the process consultation model and practitioners tended to adhere strictly to its tenets. And while this model has obvious benefits, one can likewise argue certain merits when using the expert model. This paper proposes a hybrid technique based upon a successful small group intervention. Potential problems and benefits for future application are discussed.

Process consultation (PC), as a form of third party organizational change intervention, has a number of advantages. For example, client identification of processes needing improvement, development of client diagnostic skills, and active client participation in generating remedies are but a few advantages found in the literature. (Schein, 1969; French and Bell, 1969; Huse, 1980). And although PC has long been used by consultants, it lacks a precise, operational description. Nonetheless, once consultants develop a free flowing and functional style of PC they often become locked into it and thus lose some flexibility. If an alternative consultation strategy can lead to successful outcomes, then rigid use of the PC model places a burden on both the client and the consultant since PC can consume more time than some other methods of intervention.

Are there occasions when the PC model should be modified? Can a consultant revert to "the expert mode" or the "doctor-patient model" as Schein (1969) refers to it, and still obtain successful outcomes, or should we expect the resistance from the client in accepting either the "diagnosis" or "prescription" as Schein discussed (1969, pp 6-7)? Is it not possible that in some OD efforts the PC model might reach impassable obstacles in key areas? More pragmatically, is "pure" PC (devoid of task relevant expert input) worth the time and effort in all cases? Although not intended to provide universal answers to these questions, this paper will describe an actual intervention where an alternative to pure PC was employed. It is intended to stimulate thought, to suggest application by practitioners and possibly to provoke further research.

Background

Several years ago the United States Air Force began a job redesign program using internal consultants as the change agents. The conceptual foundation for the consultants' approach to job redesign was the Job Characteristics Model (Hackman, Oldham, 1980) depicted in Figure 1. The diagnostic instrument used was the Job Attitude Survey (JAS) (Umstot, Rosenbach, 1978) which is a modified version of the Job Diagnostic Survey (Hackman, Oldham, 1975). Since there are a number of articles dealing with this conceptual framework for job redesign, no discussion will be included here.

Insert Figure 1 about here

However, the method of intervention is of concern. A PC model was developed and used for several reasons. A primary consideration was to avoid pitfalls and shortcomings of other intervention strategies as described by Schein (1969) as noted earlier. In addition, it had been the consultants' experience that long term results such as improved climates and communication, altered values, and skill development resulted from PC. Secondly, because of the wide variety of jobs inherent in an organization the size of the U.S. Air Force, it would be impossible for any one consultant to function as an "expert" for all tasks. Finally, PC can be a useful tool to ensure commitment and interventions which were not based upon a sincere desire to improve employee jobs have been observed and these uncommitted requests were generally doomed from the outset.

The particular model of PC used by this author is illustrated in Figure 2. This model depicts the development of a client-consultant relationship followed by diagnosis, client feedback and intervention. Implementation by

the client group precedes final assessment, evaluation, and termination. The joint client-consultant relationship is apparent throughout. It should be noted that during the brainstorming session, the client workgroup is tasked to generate ideas to change or redesign their job based upon the previous concepts training. The implementing group then must evaluate the recommended changes and develop action plans.

Insert Figure about here

A Modified Change Effort

The specific case described here involved a patient records section in a military hospital. At the request of the hospital registrar, the consultants visited the section, interviewed the employees, observed the work process and administered the JAS. The results of the initial survey administration are presented in Table 1. The extremely low scores accurately represent the job as observed by the consultants. Two items should be noted. The low score in skill variety reflects the limited duties: pulling records, filing lab slips in the the records, and returning the records to the file. Secondly, the Motivating Potential Score (MPS), a single index reflecting the degree to which the job will generate internal work motivation, is approximately 75 points and nearly 1.5 standard deviations below normative clerical scores (Hackman, Oldham 1980).

It should be noted that scores for climates, perceived productivity measures, and satisfactions were generally low. The nature of the task combined with the general physical environment of the section resulted in the descriptive nickname given the job by the employees --- "the Pits". The consultants privately agreed that it was an accurate description.

Insert Table 1 about here

There were arguments both for and against an intervention in this work area. On the negative side was the limited opportunity for changing the job, e.g., the technological constraints. It first appeared that little could be done to radically redesign the work without de-enriching some other jobs. When coupled with the small number of employees (N=8), it seemed that any potential gain would be small indeed for the man-hours expended. On the other hand, the high Growth Need Strength score of the employees coupled with the consultants' perceptions of strong client commitment were encouraging. In the final analysis, it was the fact that almost anything would improve the job that led the consultants to recommend continuation of the project to the client.

As is often the case in consulting efforts, changes result from two sources; (1) as an a priori process modification by the consultant or (2) as a reaction to unexpected events during the planned change process. The latter provided the impetus in this case. Specifically, the brainstorming session, which was being conducted by an intern consultant, was a nearly total failure. This was in part due to the consultant's lack of experience but was, to a greater extent, due to the inexperience of the employees. They were all young, and relatively new on the job and had little if any prior work experience. In short, they knew they didn't

like the job but could see no way to change it. As a result, instead of generating ideas for changing their jobs in the manner mentioned earlier, the employees generated and described problems. There were, in effect, no proposals for change developed in the session, even though expectations had undoubtedly been aroused by workshop participation.

This left the consultants with two alternatives: (1) declare the effort a failure beyond salvage and abandon it or (2) to abandon the normal PC model and develop "expert" solutions to employee problems since neither the hospital registrar nor any of his staff saw a solution to any of the problems. The consultants, with the approval of the registrar opted for the second alternative.

Fortunately, the administrative nature of the task was not foreign to the consultants and indeed very similar to some of the classic job redesign projects. It was therefore possible for the consultants to provide task relevant expert proposals. For example, it was recommended that the backlog of items to be filed be eliminated through a one-time effort. It was observed that although the backlog was staggering, it seemed to remain constant over time. Yet the employees perceived their efforts as contributing nothing -- "there's work when we get here and the same when we leave". Another suggested proposal was to restructure the work into identifiable units for employee responsibility, in this case by patient social security number. Prior to the intervention, quality control and error checks were made by the supervisors. The consultants recommended this function be given to the workers to increase autonomy and feedback. To increase skill variety, task identity, and feedback, it was suggested that each employee meet the patient and provide full record service to him or her. Recommendations were made to allow the employees to develop client relationships with various labs and clinics using their services.

This was congruent with another proposal permitting personnel stabilization within the section thus allowing many of the other changes to reach fruition. In total, approximately 20 possible solutions were left with the client with the clear understanding that he and his employees had the responsibility for acceptance or rejection and any implementation as outlined in the normal process model (Figure 2).

Results and Discussion

Approximately nine months after the workshops, the client requested a follow-up visit by the consultants. The JAS was readministered; the results are presented in Table 1 along with descriptive statistics. The implementation of many of the proposed changes are evidenced by statistically significant changes in three of the five job characteristics: task identity ($p=.009$), autonomy ($p=.039$) and feedback from the job ($p=.011$). As predicted by the limiting technological constraints, skill variety did not significantly increase but improvement from 1.33 to 2.29 ($p=.122$) was heartening. Most encouraging was the increase in MPS from 32 to 116 ($p=.033$).

While both psychological and communication climate improved, only the change in the former was significant ($p=.004$). All four perceptual measures of performance (based on Mott 1972), showed significant increases: quantity produced ($p=.000$), quality produced ($p=.010$), production efficiency ($p=.000$), and overall productivity ($p=.000$). While social satisfaction, which was initially high, only increased

slightly, both supervisory satisfaction ($p=.036$) and growth satisfaction ($p=.042$) increased substantially. As expected General Job Satisfaction showed a significant increase ($p=.000$). A written evaluation prepared by the hospital registrar supported the data in terms of increased morale and improved quantity and quality of output. The improved quality was later verified and noted during an independent inspection conducted by an outside agency. As expected, not all proposals for change were implemented. Approximately 50% of the recommendations were instituted in one form or another. Nonetheless, in terms of the measures described above, the intervention was an overwhelming success.

What of the departure from the pure PC model? The diagnosis of the problems was a joint effort, but clearly the treatment was in the "doctor-patient model" (Schein, 1959, p 6). Why did the expert input meet with success in this case? In all likelihood, there are several contributing yet confounding factors. Certainly the positive elements in existence prior to the project contributed, e.g., high employee growth need strength and strong client commitment to improvement. But perhaps there are more fundamental questions concerning the PC model. For example, how much client ownership comes from each phase of the project? Certainly some comes from the client's knowledge that he or she can terminate the project at a number of points in the consultation process. Some ownership comes simply from the act of participating in the workshop; however, the consultants had assumed that a great deal of ownership emerged from employees knowing they had developed the changes to their jobs. That assumption can certainly be questioned in this case. Perhaps the real ownership in the change process comes from the experience of being able to participate in the implementation of change proposals. In the present case, after proposals from the brainstorming session had been evaluated, members of the client work group were very open to the consultants' suggestions for changing their jobs. In interviews conducted immediately following the intervention as well as nine months later, there were no perceptions of consultant manipulation by either the employees or the supervisors. No derogatory mention of the consultants' expert input was made in the client's written evaluation, discussed earlier.

Conclusion

Whether this project is viewed in terms of the changes in scores of the measured variables or the client evaluation, it was a successful intervention. And certainly, the expert mode was used in a critical portion of an otherwise accepted PC model. The expert mode was successfully employed in this situation but the conditions necessary for its general application remain unclear.

Certainly the intervention described here should not be construed to devalue either the PC model or its underlying philosophy.

More than anything else, the present case demonstrates the need for a normative model for application of a combination of the expert and the pure process consultation model. A great practical benefit would be the savings of valuable time on the consultant's and particularly the client's part. And if used effectively, in a broader sense, it would permit the experience of the knowledgeable consultant to be injected more freely into many situations where the client is at a loss for alternatives. The results of additional research have implications for expanding our knowledge of the consulting process as well as for practice.

References

- French, W., and Bell, C. Organization Development: Behavioral Science Interventions for Organization Improvement. Englewood Cliffs, N. J.: Prentice-Hall, 1969.
- Hackman, J.R., and Oldham, G.R. "Development of the Job Diagnostic Survey", Journal of Applied Psychology, Vol. 60, 1975, pp. 159-170.
- Hackman, J.R., and Oldham, G.R. Work Redesign. Reading, Mass.: Addison-Wesley, 1980.
- Huse, E.F. Organization Development and Change. St. Paul: West, 2nd ed., 1980.
- Mott, P.E. The Characteristics of Effective Organizations. New York: Harper & Row, 1972.
- Schein, E.H. Process Consultation: Its Role in Organization Development. Reading, Mass.: Addison-Wesley, 1969.
- Umstot, D.D. and Rosenbach, W.E. "Job Attitude Survey", United States Air Force Academy, Colorado, 1978.

Table 1

Descriptive Statistics for Job Attitude Survey Results

<u>VARIABLES</u>	<u>Pre-test</u>	<u>Post-test</u>	<u>t</u>	<u>Prob.</u>
	N=8			
Skill Variety	1.33	2.29	1.66	.122
Task Identity	3.11	5.13	3.13	.009
Task Significance	5.44	5.79	.56	.587
Autonomy	2.94	4.63	2.31	.039
Feedback from Job	2.83	5.08	2.99	.011
MPS	32.33	116.26	2.41	.033
Psychological Climate	3.02	4.83	3.53	.004
Communications Climate	5.17	6.29	2.10	.058
Quantity Produced	3.00	4.25	6.55	.000
Quality Produced	3.50	4.37	3.06	.010
Production Efficiency	2.33	4.13	5.60	.000
Overall Productivity	2.94	4.25	6.64	.000
Social Satisfaction	5.11	5.92	2.02	.066
Supervisory Satisfaction	4.61	6.25	2.36	.036
Growth Satisfaction	2.67	4.54	2.27	.042
General Job Satisfaction	1.94	4.92	4.90	.000
Growth Need Strength	6.42	6.46	.08	.900

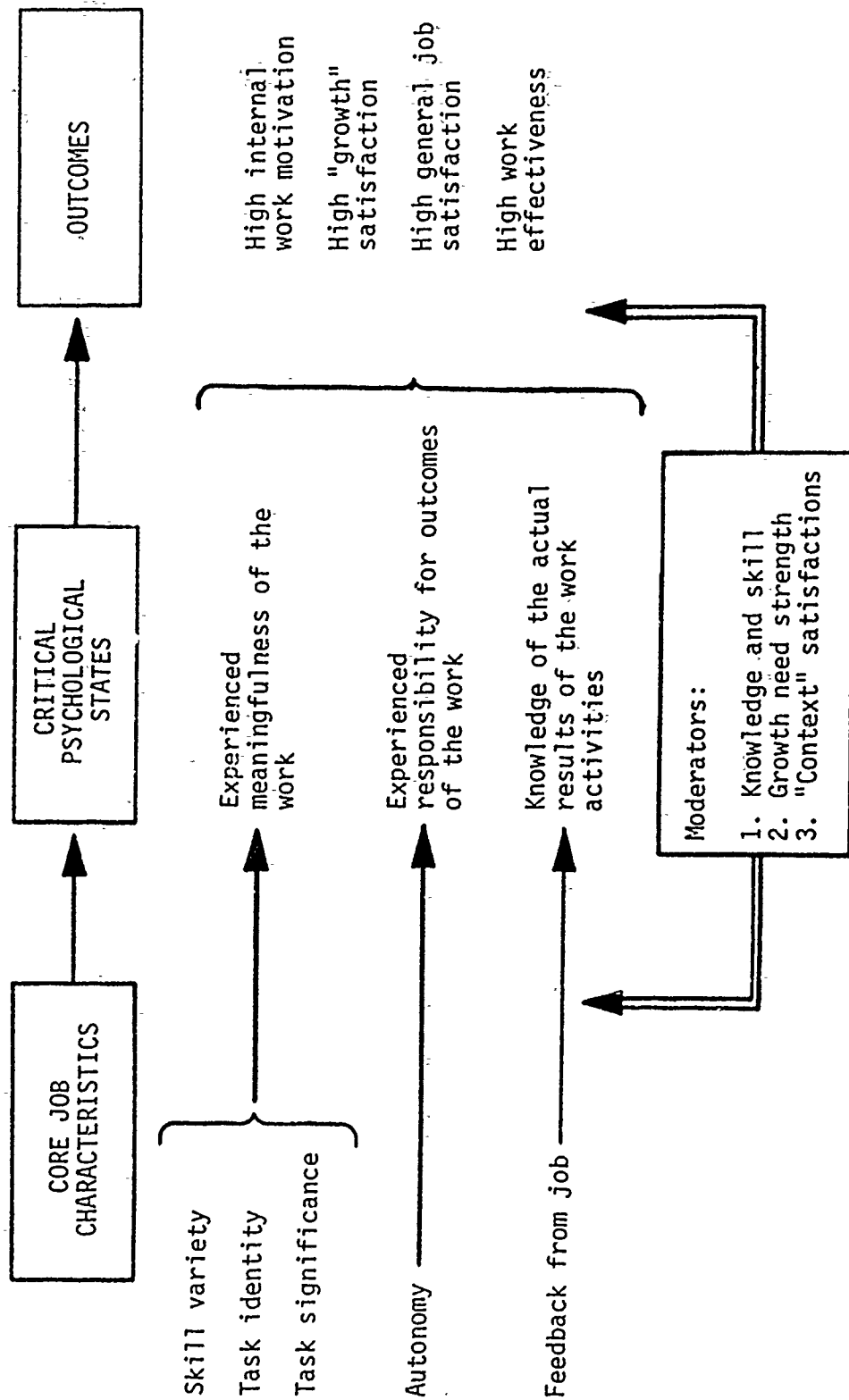


Figure 1

The Job Characteristics Model

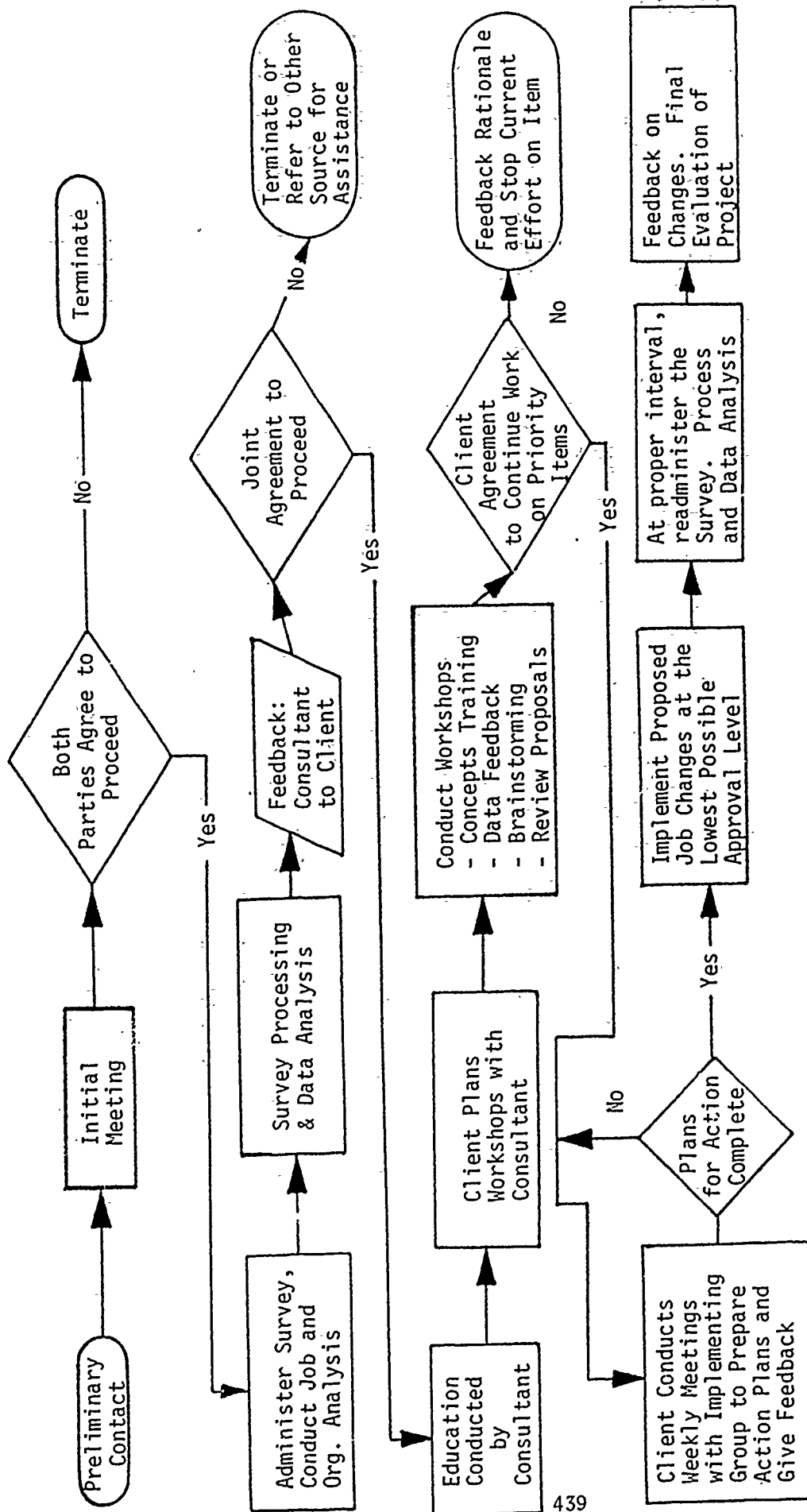


Figure 2
Process Consultation Model for Job Redesign

INFLUENTIAL ASSOCIATIONS IN ORGANIZATIONS:
THE 'TOUCHSTONES' OF CAREER SUCCESS

Frank R. Wood
Department of Behavioral Sciences and Leadership
USAF Academy

Rosanna Hertz
Department of Sociology
Northwestern University

Abstract

Role models, advisors, mentors, and sponsors are described as qualitatively different types of influential associations in organizations.

Individuals in organizations find the navigation of careers a difficult task. This task is made easier by certain influential associations which are the "touchstones" of career success.

This paper will delineate the types of influential associations normally found in organizations, namely: role models, advisors, mentors, and sponsors. While each of these has been used interchangeably by previous researchers, we propose they are qualitatively different categories of relationships that have differential impact on individual career paths in organizations. In this paper, we prefer not to enter into the standard debate of whether these associations generate career success. Instead, we will delineate categorical differences that should sensitize future research in this area. A thorough understanding of how these informal mechanisms operate is an important step toward understanding why organizations which structure for equality at the "starting gates" still generate differential career outcomes for men and women.

METHODS

This paper links the findings uncovered by the authors in their individual studies of two very different organizational settings, namely, military and corporate organizations. While neither study was originally designed to study these specific issues, both asked questions relating to individual career trajectories and the importance of others for career progression. In both studies, the data collected showed influential associations to be important, but each study only revealed part of the picture. Combined, they illustrate how these significant

others "really operate" to influence an individual's career.

Each study employed in-depth interviewing with respondents who had established careers in the traditional sense, that is, they had organizationally established opportunities for advancement. One study, by Wood (1980), included 83 Air Force junior officers, in their first ten years of service and the other study, by Hertz (1980), included 60 individuals established in corporations at the middle to top levels of management. Both groups provide strikingly similar accounts of the phenomenon being examined. Thus, the data will be reported as a whole. Further, since this paper represents a theoretical rethinking of the problem, our data will be used primarily to illustrate and support the categories we suggest.

INFLUENTIAL ASSOCIATIONS

Significant associations in organizations can be divided into four distinct categories: role models, advisors, mentors, and sponsors. The specific dimensions of each are shown in Table 1. The reason that these associations are informally important in organizations is that they have the potential for facilitating movement through the organization. At critical junctions these associations may well mean the difference between not only success and failure, but also the difference between rapid advancement and minimal or slower movement.

TABLE 1

INFLUENTIAL ASSOCIATIONS

	ROLE MODEL	ADVISOR	MENTOR	SPONSOR
	-----	-----	-----	-----
RELATIONSHIP BASIS	shared problems	knowledge	wisdom	power
POWER LEVEL	peer or higher	peer or higher	lower to higher	higher only
INTERACTION	one-way	one-way	exchange	exchange
INITIATED BY	individual	individual	mentor	sponsor
COMMITMENT	none	limited	personal	instrumental

ROLE MODELS

Perhaps the most talked about category of influential associations in the literature is role models. These individuals have been characterized as either being embodied in a single individual, fictitious or real (for example, Don Quixote) or as a theoretical construct involving attributes of several persons. (Bucher and Stelling, 1977) Role models are selected by the individual and may be characterized as a one-way relationship with little or no interaction. Usually, they are exemplars which represent the solution to problems as yet unsolved by the individual. These may be specific problems such as the presentation of oneself in a work situation or more holistic problems such as how to integrate conflicting aspects of work and family life. Simply put, role models are selected by the individual for having achieved what the individual wishes to accomplish. For example, one woman in our sample who was trying to solve a family/work dilemma described how she used role models found in friends who had already been in her situation:

My friend, who works in New York, she and her husband went to business school at the same time my husband and I did -- they have a young child. I don't know how she does it, but... She is kind of a role model for me.

The role model below illustrates how composites of individual traits are constructed:

There are two people that I admire and respect. One is my immediate supervisor. He shows genuine interest in his job, sets priorities, and gets the job done. He values my opinion. The other guy is the division chief. He is a good worker and a thinker. He explains "why" the Air Force does certain things.

ADVISORS

Advisors disseminate specific information and facilitate the solution of specific problems. Formally, they can be provided by the organization as organizationally defined positions or as departments (e.g., graduate advisors, career counselors, personnel departments, tax advisors, etc.) Informally, they may be selected by the individual as with role models, that is, an individual may approach another who has mastered a specific expertise or solved a particular problem. Like role models, the contact is initiated by the individual and while there is interaction, it is specific and limited to the problem at hand. Unlike the role model, where the basis of the relationship is some shared problem, the basis of the advisor-advisee relationship is either greater knowledge or proven expertise. Advisor relationships, then, are segmented relationships of short duration. This respondent illustrates the greater importance placed on informal advisors over those formally established:

If you want to know about your career, you have to go to the Central Base Personnel Office and they don't know what they are doing. I would rather have someone in my organization tell me the advantages and disadvantages of a career move.

MENTORS

Mentors have been more recently reported by many researchers as critical for career advancement. These people have documented the scramble of younger people to find one, the older people to be one, and the universal belief that career success is predicated on participation in this type of relationship. The mentor is important because of his greater experience and concern for the individual. Essentially, he provides wisdom and support. In doing so, he encourages the individual in directions that are right for the apprentice and not necessarily the organization. The wisdom of the mentor is based on his own personal successes and failures and through the relationship, the personal histories of each are wedded. While a mentor relationship may begin when both are members of the same organization, it may continue after either individual has severed the initial organizational affiliation. Unlike role models and advisers, the mentor relationship is an exchange relationship. The mentor provides support for the apprentice's needs and, in turn, receives a sense of personal regeneration. (Levinson, 1978; Erikson, 1968) Of critical importance is that this association is established by the mentor and is characterized by a strong sense of personal loyalty. For example, one woman who was at the early stages of her career described her relationship with a mentor as follows:

There's a man here in this office that I think has done a lot for helping me see my way through and my being able to accept responsibility for managing other people.

The mentor's concern for the individual is illustrated by the following description:

He listens to me and understands how I feel. He empathizes with me. He seems to have a feeling of concern for what I am going to do with my life. He is concerned for my welfare as much as the Air Force.

Interestingly, a mentor, as the quote below illustrates, need not be someone senior in rank, only in experience. As one military officer commented about a subordinate:

It was a Chief Master Sergeant who taught me how to be an officer and gave me a sense of purpose and direction.

SPONSORS

A sponsor relationship is an instrumental relationship which involves the use of power. Sponsors select "informal assistants" to help them in their jobs. (Graen, 1977) In return, they reward these assistants by employing their power to give them special treatment such as arranging for them to gain the experiences necessary for their advancement. (Becker and Strauss, 1956) Like mentors, this is an exchange relationship; however, at the core of this association, is a concern for joint career mobility. Therefore, this relationship is instrumental and lasts only insofar as it is beneficial to both the assistant and the sponsor. For example, one assistant, whose corporation was acquired by another corporation, was given the option of moving with her sponsor. As she explained:

The man that I worked directly for was the Vice-president of Corporate Planning and Development. He and the President would be moving to the new corporation headquarters. I had a number of options that I talked about with my boss: they would take me with them and I would be given a higher level staff job or I could stay put and take over the planning function with the new people.

The effect of a sponsor's power can influence a number of job aspects as reported by this respondent:

My first squadron commander made sure I got good evaluations, sent me on trips which provided good experience, and gave me a good job with visibility.

PREVIOUS DIFFICULTIES WITH DELINEATION

In previous research, one or more of these categories have been assumed to stand for the rest or all have been used interchangeably, e.g., Levinson (1978:24) describes a "true mentor" as a teacher, advisor, sponsor, friend, relative, neighbor, or exemplar. This confusion of terms probably arises because in reality, one or several of the functions described may be embodied in the same person. For example, a person who provides the social support associated with mentoring may also offer an exemplar which solves the dilemma of childcare arrangements, normally a function of role models. Further, if a person has the organizational power associated with a sponsor, he or she may also be able to make special arrangements to accommodate immediate problems so that the individual's future potential is not lost to the organization.

By reconceptualizing these relationships, we are suggesting that researchers focus on the process or functions represented by these categories and not on the persons who may be linked with them. Only through such delineation can we gain an understanding of how these influential associations really operate in organizations.

REFERENCES

- Becker, H.S. and A.L. Strauss
1956 "Careers, Personality and Adult Socialization,"
American Journal of Sociology, Vol. 62, No. 3, pp. 253-266.
- Bucher, R. and J.G. Stelling
1977 Becoming Professional. Beverly Hills: Sage.
- Erikson, E.H.
1968 Identity: Youth and Crisis. New York: Norton
- Graen, G.
1977 "Role-Making Processes Within Complex Organizations,"
in M. Dunnet (ed.), Handbook of Individual and Organ-
izational Psychology. Chicago: Rand McNally.
- Hertz, R.
1980 "Dual-Career Couples in the Corporate World,"
Dissertation, Northwestern University, Evanston, Ill.
- Levinson, D.J.
1978 "Growing Up With the Dream," Psychology Today, Jan.
- Wood, F.R.
1980 "USAF Junior Officers: Professional Identity
and Commitment," Dissertation, Northwestern University
Evanston, Ill.

Values and the Foreign Military Officer

Lynville E. Taber
Air Command and Staff College

ABSTRACT

Each year numerous military officers from foreign nations enter our military training and educational institutions. This study was designed to determine if their values change as a result of this experience. A significant change was noted on only one of six values measured.

Each year numerous military officers from foreign countries travel to the United States to receive training or formal education at United States military installations and civilian institutions. The training and/or education of personnel sought by the leadership of the foreign countries varies from a few weeks of highly technical schooling offered at military installations throughout the United States to several months of more formal instruction at professional military education (PME) courses, e.g., National War College, Air War College, Naval War College, Command and General Staff College, Air Command and Staff College, Naval Post-Graduate School, Squadron Officer School.

One such PME course is conducted each year by the United States Air Force at Air University, Maxwell Air Force Base, Alabama. This course, Air Command and Staff College (ACSC), is designed to provide instruction to the United States military officer at the mid-point in his or her military career, i.e., ten to thirteen years of commissioned military service and holding the rank of major.

Since 1954 over 5,400 foreign officers from 77 countries have completed the Air University's Foreign Officer School which prepares them for entry into one of the University's Professional Military Education (PME) courses (Air University, 1980). Many of these officers have then entered ACSC to pursue a program of study from August through February of each year.

While at ACSC the foreign student officers undergo three phases of curriculum, one of which is called "Command, Leadership, and Resource Management." The goal of this phase of the curriculum is to have student officers organize their personal and professional knowledge, attitudes, values, and skills into a personal credo for military leadership. The thrust of the curriculum design, especially the values area, encouraged this writer to investigate whether the course work presented had any significant impact upon the values held by the foreign student officers.

Defining what is incorporated in the term "values" has been a subject of controversy for a number of years, yet the term values is used freely, and the general public seems to know what is meant by a "value." Kluckhohn (1951) defines a value as a conception, explicit or implicit, distinctive of an individual or characteristic of a group, of the desirable which influences the selection from available modes, means, and ends to action. Roscoe (1968) defines values as personal and social ideals, attitudes, beliefs or standards which may be used to evaluate and regulate the actions of the individual and society. Foss (1977) proposes that values are beliefs about classes of objects, situations, actions, and wholes composed of them in regards to the extent that they are good, right, obligatory, or ought to be. A more general definition is offered by Mankoff (1974) who concludes that it is the basic value system to which a person subscribes that ultimately determines who he is, what he is, and how he relates to himself, his family, other people, his job, his boss, indeed, the whole world around him. Spranger's (1928) work defends the view that personalities of men are best known through a study of their values or evaluative attitudes. It is from this conceptual framework that Allport, Vernon and Lindzey (1960) developed their instrument for studying values. Spranger's thesis is that there are six basic interests or motives in personality: the theoretical, economic, aesthetic, social, political, and religious.

The study of values held by members of the military service is not a new development, but since the Vietnam era research has increased considerably in terms of gathering data on United States servicemen. No research was evident in terms of the values held by foreign military officers.

Janowitz (1971) and Moskos (1976, 1977, 1978) share the opinion that because of the changing social structure and the new values of civilian society in general since the mid-1960's, the military service has a new generation of recruits and officers entering the armed forces, holding much different standards of personal behavior, yet they are but a reflection of the society from which they came. Studies conducted by both military and civilian researchers have determined that military enlisted men and officers do not differ significantly from their civilian counterparts whether they enter service through basic training courses or through a service academy. (Hecox, 1970; Neuman, Abrahams & Githens, 1972; Bachman, 1974; ACSC, 1978; Boyle & McCall, 1976; Short, Westen & Jones, 1978)

A number of researchers share Power's (1980) opinion that groups of individuals generate shared norms, values and attitudes, which are experienced as properties of the group itself. That is, once an individual has entered the military service, he or she begins to adopt the values or adapt to the values of the organization. If such an adaptation cannot take place, the individual will exit the organization at the earliest opportunity (Lovell, 1964; Wemuth, 1970; Janowitz & Little, 1974; GAO, 1976; Hudak, 1976; Goodpaster, 1979).

It has been a "given" for some time that values play a major role in motivation toward a particular occupation. In the civil sector as well as in a purely military environment, if an individual is able to fulfill certain of his/her values on the job, then a motivation toward that job or a certain kind of "job satisfaction" results (Gordon &

Medland, 1964; Githens, 1966; Walls & Gulkus, 1974; Brown, 1976; Ronen, 1978, Kidron, 1978). Coffey and Apply (1968) do not agree with this assertion however and conclude that the relative contribution of values to motivational factors has not been adequately assessed and that its conceptual status is uncertain at this time.

The persistence of values has been an item of research for some time also; that is, can a person's value set be changed as a result of education? Massey (1975) believes that it may be possible in that values are "implanted" during the first twenty years of life; after that, however, one must experience a "significant emotional event" in order to bring about change. Other researchers believe that not only can values be changed by experiencing an educational environment, the direction of change can be predicted by a particular kind of college education (Newcomb, 1943; Newcomb & Feldman, 1968; Kaats, 1969). Yet, considerable disagreement exists in the current literature. Kuh (1976) found that values held during college persisted for five years beyond graduation, and there is strong evidence to support the contention that values and attitudes held upon entering college will persist ten or more years beyond graduation (Kirchner & Hogen, 1972; Jansen, 1974; Schein, 1975; Mortimer & Lorence, 1979).

The purpose of this study was to examine whether a values change occurs in foreign military officers attending ACSC for nearly six months.

METHOD

Subjects

The subjects for this study were 42 foreign military officers in their respective Air Forces' from 27 countries throughout the world. The student officers were male with an age range from 26 to 51; mean age, 36.3 years old. Their military rank spread from major to colonel, with 99% at the rank of major. Twenty-seven of the foreign officers were pilots and the remainder were spread evenly amongst nine other occupational specialties, i.e., personnel, logistics, intelligence, engineering, etc. The reported religious preference for the officers follow: Agnostic-2; Buddhism-5; Catholic-11; Hebraism-1; Islam-16; Greek Orthodox-1; Protestant-4; Other-2.

Instrument

The values held by the foreign military officers were measured using the Study of Values (Allport, Vernon & Lindsey, 1960). This is a 45-item test, 30 of which are forced choice, and 15 of which require the rank ordering of four alternatives. The Study of Values aims to measure the relative prominence of six basic interests or motives in personality: the theoretical, economic, aesthetic, social, political, and religious. The classification for these value sets are based upon Spranger's Types of Men (1928).

Norms for the instrument are based on 1,816 college students. Split-half reliabilities (N=100) show Spearman-Brown product moment correlations ranging from .84 (theoretical) to .95 (religious). A Z transformation provides a mean reliability coefficient of .90. A final item analysis

(N=780 male and female subjects at six different colleges) indicates that every item is positively correlated at the .01 level with the total score for its value. Test-retest reliabilities after one month (N=34, 1951) and after two months (N=53, 1957) range from .84 to .93 for the economic and religious scales, respectively, for the 1957 sample. For the 1951 study, .89 was the Z transformed mean repeat reliability coefficient; for the 1957 study, this figure was .88 (Allport, et al., 1960; Robinson & Shaver, 1973).

Procedure

On August 21, 1980, each subject was asked to complete the Study of Values in class. Five and one-half months later, February 5, 1981, a posttest was administered, again, the classroom setting. N=42.

Results

Using the t -test, only one of the six variables showed significant differences as a result of the posttest. The Social Scale (pretest $M=39.00$, posttest $M=35.83$) showed a decrease of significance (Mean difference, -3.167 , $t = -3.87$, $p < .0001$). (See Table 1.)

Table 1

Means, Standard Deviations, Mean Differences, and t Values for Pretest and Posttest Measures, the Study of Values

<u>Variable</u>	<u>Pretest</u>		<u>Posttest</u>		<u>Mean Diff</u>	<u>t</u>
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>		
Theoretical	44.19	5.76	44.71	6.14	+0.52	0.57
Economic	42.88	6.25	44.14	8.22	+1.26	1.50
Aesthetic	35.43	7.45	34.36	6.65	-1.07	0.95
Social	39.00	5.56	35.83	6.18	-3.17	3.87 *
Political	43.79	6.12	45.64	6.56	+1.86	1.62
Religious	34.90	8.14	35.21	7.92	+0.31	0.30

* $p < .0001$

Discussion

The findings support, in part, the contention of many researchers that values will not change short of a significant emotional event in a person's life. The fact that ACSC is an experience of a rather short duration for the foreign military officer, taken in combination with his total exposure to the American culture during his stay in the United States, one would not expect too drastic a change in the individual's values profile.

Because this sample represented a melting pot of the world (27 countries from the four corners of the globe), it is possible that a "leveling" effect occurred in terms of allowing the instrument to detect any significant shifts in values held. The fact that one value criteria changed significantly may simply reflect that four others changed slightly and a trade-off resulted.

These findings warrant further substantiation before assuming generalizability to other classes for foreign military officers at ACSC or other United States courses, military or civilian.

BIBLIOGRAPHY

References

- ACSC-78, basic airmen's values and perceptions of society, Readings in Command and Management, 1978, 5, 187-193, Maxwell AFB, Alabama: Air Command and Staff College.
- Air University Catalog, Air University, Maxwell AFB, Alabama: 1980-1981, pp. 37-38.
- Allport, G. W., Vernon, P. E., & Lindzey, G. A Study of Values, (3rd Ed) Boston University, 1960.
- Bachman, J. E. Values, preference and perception concerning military service, University of Michigan Institute for Social Research, Part II, February 1974, pp. 1-25.
- Boyle, P. J., and McCall, M. P. The value hierarchies of selected Air Force Academy classes, USAF Institute of Technology Report, Sep 1976, p. 2.
- Brown, M. A. Values - a necessary but neglected ingredient of motivation on the job, Academy of Management Review, October 1976, p. 22.
- Coffey, C. N. and Appley, N. H. Motivation Theory and Research, New York: Wiley and Sons, 1968.
- GAO-76, Student attrition at the five federal service academies, Comptroller General of the United States, March 1976.
- Foss, D. C. Value controversy in sociology, San Francisco: Jossey-Bass, Inc., 1977.
- Githens, W. H. The values of junior officers, obtainability and comparability of various career values, Navy Personnel Research Activity, January 1966, p. 7.
- Goodpaste, A. J. Moral choices, military ethics and national values in the 1980's, Journal of Liberal Education, 1979, 65, 249-258.
- Gordon, L. V. and medland, F. F. Values associated with military career motivation, U.S. Army Personnel Research Office, March 1964, p. 5.
- Hacox, W. E. A comparision of new cadets at USMA with entering freshman at other colleges, class of 1973, Office of Institutional Research, U.S. Military Academy, February 1970, p. 24.
- Hudak, S. J. The psychological acclimation of service academy selectees (Doctoral dissertation), Journal of Higher Education, 1976, 1428A.

- Janowitz, M. and Little, R. Sociology and the Military Establishment, Beverly Hills, California; Sage Publications, 1974.
- Janowitz, M. The Professional Soldier (2nd Ed). New York: McMillan, 1971.
- Jansen, D. and Ban, E. Values and attitudes: change from grade ten to masters, American Personnel and Guidance Journal, 1974, 2, 82-90.
- Kaats, G. Developmental changes in belief systems during a service academy education, Proceedings of the 77th Annual Convention, American Psychological Association, 1969, 4, 651-652.
- Kidron, A. Work values and organizational commitment, Academy of Management Journal, 1978, 21, 239-247.
- Kirchner, J. and Hogan, R. Student values: a longitudinal study, Psychology, 1972, 9, 36-39.
- Kluckhohn, C. Values and value orientation in the theory of action: an exploration in definition and classification, in T. Parsons & E. Shils (Eds) Toward a General Theory of Action. Cambridge, Massachusetts: Harvard University Press, 1951.
- Kuh, G. Persistence of the impact of college on attitudes and values, Journal of College Student Personnel, 1976, 17, 116-122.
- Lovell, J. The professional socialization of the West Point cadet, in Janowitz M. (Ed) The New Military: Changing Patterns of Organization. New York: Wiley & Sons, 1964.
- Massey, M. What you are in where you were when, A film produced by the Magnetic Video Library, Farmington Hills, Michigan, 1975.
- Mankoff, A. Values - not attitudes - are the real key to motivation, Management Review, 1974, 23-29.
- Mortimer, J. and Lorence, J. Work experience and occupational value socialization: a longitudinal study, American Journal of Sociology, 1979, 84, 1361-1385.
- Moskos, C. Jr., Armed forces in american society: from institution to occupation, Keynote address given at the Fifth Psychology in the Department of Defense Symposium, U.S. Air Force Academy, Colorado, February 1976.
- From institution to occupation: trends in military organization, Armed Forces and Society, 1977, 4, 41-50.
- Military compensation and the military institution, Unpublished paper, delivered at a symposium, USAF Academy, Colorado, 1978.

- The all-volunteer military: calling, profession or occupation, Parameter, Journal of the US Army War College, 1976, 1.
- Newcomb, T. Personality and social change: attitude formation in a student community. New York: Dryden Press, 1943.
- Newcomb, T. and Feldman, K. The impacts of colleges upon their students. Ann Arbor, Michigan, 1968.
- Neuman, I., Abrahams, N., and Githens, W. The values of junior officers: the relationship between career values and retention, Naval Personnel and Training Research Laboratory, San Diego, 1972, Report SSR 72-23.
- Power, C. Further reflections on values assessment, Personnel and Guidance Journal, 1980, 58, 616-617.
- Robinson, J. and Shaver, P. Measures of social psychological attitudes. Ann Arbor, Michigan, Institute for Social Research, 1973.
- Ronen, S. Personal values: a basis for work motivation, satisfaction and work attitude, Organizational Behavior and Human Performance, 1978, 21, 80-107.
- Roscoe, J. Ritter, C., Eglovic, S., and Thayer, J. American college student values: Preliminary report of a nationwide survey, Colorado Journal of Educational Research, 1968, 8, 3-26.
- Schein, E. The stability of values in the first ten years of the career, Office of Naval Research Reports, 1975, (ONR #N00014-67-A-020400073).
- Short, L., Westen, R., and Jones, G. Values of Air Force Academy cadets: a comparative study, Unpublished paper, USAF Academy, April 1978.
- Spranger, E. Types of men. Translated from the 5th German edition of Lebenformen, by P. J. Pigors. Halle: Max Niemeyer, Verlag, 1928, American Agent: Stechert-Hafner, Inc., New York.
- Walls, R. and Gulkus S. Reinforces, values, and vocational maturity in adults, Journal of Vocational Behavior, 1974, 4, 325-331.
- Wermuth, A. The impact of changing values on military organization and personnel, Advanced Studies Group, Waltham, Massachusetts, 1970, p. 29.

A COMPARATIVE STUDY OF VALUES OF MILITARY MEN AND WOMEN

by

Jose Bolton
Chief of Social Actions
RAF Mildenhall, APO NY 09127

THE PROBLEM: The problem of the study was to examine selected values of male and female military personnel to determine if there were psychosocial determinants which Air Force managers could find useful in terms of recruitment. With an all-volunteer service and increased utilization of women in 95 percent of all Air Force career fields, it seemed that necessary policy changes, which must follow, should be made with the best possible comprehension of what will be different personnel versus system needs in the future.

The objective of the study was to determine, through a study of values questionnaire, the differences and similarities of selected values between males and females when compared by time in service and sex. For the present study, groupings were made by one year or less in service and four years or more in service.

The importance of the study was derived from the fact that the numbers of women entering the Air Force are on the increase and, by 1984, projected to be approximately 80,000, representing 15 to 16 percent of the total force strength. The Air Force will spend hundreds of thousands of dollars for educating, training and equipping its members. The greater numbers of women will require that these areas be modified in order to avoid the loss of valuable human resources, both male and female. It is presumed that the results of the study will provide the Air Force manager with data to predict Air Force needs and establish policies which will aggressively pursue that end.

METHOD. All data were secured from voluntary responses on the questionnaire employed to gather data from 200 participants. Data were grouped by time in service and sex, reported by group means and analyzed using the Student's t test for significance at the .05 level.

RESULTS. The results of the study indicated that scores for all males and all females on the Rotter I-E were significantly different. Males were more internal than were females. This was true for all groups of males and females regardless of time in service. Time was a factor for women when analyzing internal movement on the I-E. Women became more internal with the increasing length of time in service. Men also became more internal with the increase of time in service, but not statistically significantly so. As women entered the Air Force, there appeared to be an area of difference between males and females which was identified by the I-E, but this difference was reduced with the increase of time in service.

A COMPARATIVE STUDY OF VALUES OF MILITARY MEN AND WOMEN

by

Jose Bolton, Ph.D.

In 1972 the Social Actions program became a very unique program in the annals of military history. It was created to assist the chain of command in dealing with the numerous social pressures which surfaced in the United States in general and the military in particular. Many people believed that if individuals became aware of these social problems, there would be fewer problems.

Although Social Actions was composed of three separate functions (Drug/Alcohol, Equal Opportunity and Treatment, and Human Relations Education), Human Relations Education (HRE), initially called Race Relations, received the greatest emphasis. In 1975 discussions of women in the Air Force became part of the HRE curriculum. Intituled "Utilization of Women In The Military" and it was designed to deal with the psychosocial adjustments experienced by Air Force men and women. I felt these added adjustments by men and women were influenced by two major factors: First, the "All-Volunteer Force" which combined with the national economic picture at the time (1975) to bring more women into the Air Force. Secondly, in 1975 the Air Force opened to women, numerous career fields that were previously available only to men.

The reaction in male dominated fields to this perceived invasion of the male domain ranged from enthusiastic acceptance to overt rejection of women in the work areas. Furthermore, proportionately more women failed, were discharged, and fewer were retrained out of these nontraditional career fields than men. Thus, greater numbers of women in the Air Force and the opening of male dominant jobs made assimilation of women a crucial issue.

This study was concerned with the influence of socialization on males and females in the Air Force. What are the key factors in the assimilation of women into the Air Force? Does assimilation influence retention? Many females have been successful in the Air Force. I believe that those women who did succeed, succeeded because they successfully assimilated into the Air Force. Those who failed, failed because of unsuccessful assimilation. This paper is designed to explore the characteristics of successful assimilation techniques which the Air Force can adopt to assure a greater success rate for women,

A review of the literature shows that very little work has been done in the area of psychosocial attitudes of military men and women and how these could affect behavior, or how the attitudes generally are influenced by and in turn influence the self-concept. A person's self-concept is determined by various factors, some of which neither the individual nor society can alter; such as sex. But independent factors like education and environment, and psychological factors such as self-concept greatly influence behavior as well. An aspect of self-concept is a need to control the environment. (Lefcourt, 1972)

This study examines how values can influence the assimilation of military men and women. For the purpose of this research, "value", is defined as: The core component of a clustering of attitudes which directs behavior on a long-range basis toward some goals in preference to others. The examination of personal values can have a critical meaning for any society that is concerned with the equal treatment of all members. This paper presents data that can aid Air Force managers to understand human nature and indirectly assist in the formation of policy regarding the use of men and women in nontraditional careers,

The theoretical foundation for research on values was set by Spranger (1928), Lasswell (1959), Rucker (1978), Rotter (1966), Allport and Vernon (1931), and Allport, Vernon and Lindzey (1951). The Rotter Internal Versus External Locus of Control of Reinforcement (I-E) scale was used to gather the data. The Rotter gives an indication of the participants' perception of their influence over their environment. The I-E refers to a continuum of generalized expectancies that reinforcements are self-determined (that is internal) as opposed to being a function of luck, chance, or powerful others.

The research was conducted between October 1979 and March 1980 with the I-E being administered to 200 randomly selected military men and women. Each group was subdivided by time in service and a 2X2 factorial design was used with sex and time in service as measurement variables.

2X2 Factorial Design With Sex and Time
In Service as Measurement Variables

Time in Service	Males (N=100)	Females (N=100)
1 Year or Less in Service	50 Males	50 Females
4 Years or More in Service	50 Males	50 Females

Data was gathered which lead to mean scores that inturn were analyzed for differences using Student's t test with a .05 level of confidence, with 98 degrees of freedom. The lower the score the greater the degree of self-confidence, making adjustment to the Air Force environment easier. That is those with lower scores assimilated more easily. An examination of the mean scores of males and females on the I-E shows that males had a locus of control of 8.10, while females' locus of control after four years of service was 8.70, and females with one year or less in service was more than 9.3. Males with four years or more in service had a mean locus of control under 8.0 which was more internal than the locus of control of males with one year in service, over 8.2.

A Comparison of I-E Scores of Military Men
And Women by Sex and Time in Service

Sex	Grouping	N	X
Male	1 Year	50	8.24
	4 Years	50	7.96
Female	1 Year	50	9.36
	4 Years	50	8.70
Male	1 Year	50	8.24
Female	1 Year	50	9.36
Male	4 Years	50	7.96
Female	4 Years	50	8.70
Total	Male	100	8.10
	Female	100	9.03 *

* < .05,98

This means that males in all groupings perceived themselves in control of their lives, more so than women. But the fact is that all groups, men and women, after 4 years in the Air Force had lower scores; thus perceiving themselves to be in greater control than at their starting points. There is a statistically significant difference between males and females (8.10 and 9.03 respectively), but when comparing these groups over time in service, little difference was found. The locus of control for women continues to be more external than for men.

It is hard to say what environmental factors have affected males and females and caused this measurable difference. Using the male score as the norm, because the military is traditionally a male domain, one can quickly see that females have not been socialized in the same manner. Lewis (1972) reported that boys are encouraged to be more aggressive and more independent than are girls.

Males in most societies are required to be independent. At a young age they learn that they have to be aggressive if they want to get ahead. Females, on the other hand, were taught how to be dependent and allow others to satisfy their needs,

The difference of I-E scores for males and females is also a result of age, time in service, rank, level of authority, and maturity. All of these factors

contribute to how individuals view themselves. However, it should not be assumed that because women are performing the same duties as men, they will have the same philosophy of life (Saxton, 1977).

There are numerous conclusions to be drawn from the data I have gathered. With each conclusion there are more questions, but one factor is consistent. Conformity to society's code, the Air Force in this case, is a critical part of the change process. While there were differences in all groupings, only one was found to be statistically significant.

Time in service was a factor in changes for all groups, but significantly so between males and females in the 1 year in service group as measured on the Rotter I-E. In a related study, preliminary data I gathered on the Allport, Vernon and Kindzey suggest that women entering the military have significantly different values than a similar group of civilian women.

Little data is available at this time which could give a clear picture of the psychosocial motivators that men and women have when they come into the Air Force. However, my research shows women who had scores that were more internal frequently came from families which had working mothers or mothers as single parents. These role models may have influenced the socialization of many of the females in the study.

Men and women encounter a similar socialization process when they enter the Air Force, but since the military is a male dominant society, based on male norms and skills, it follows that male adjustment would be less difficult. This adjustment is reflected in the changes of scores over time. The mean score for males changed over 4 years, almost .3 of a point, while for females it moved nearly .7 of a point.

I believe the female score changes can be attributed to assimilation. It is apparent that when females, like any outgroup, gained knowledge of the system they felt they were better able to influence their futures. Lefcourt (1973) saw these as changes that were required by an individual's need to control his/her situation or environment. As recruits achieve, or believe that they have achieved, a greater measure of control in their environment, feelings of helplessness diminish and they begin to function more effectively in their environment.

In terms of female recruitment, it is important to consider the trend toward external scores for females in early service years. At present 6 weeks of basic training are required for all recruits, and during that time they also must learn to adjust to military life. Females make this adjustment as well as males but when they arrive at their first assignment, the adjustment is not to a training instructor but to a "REAL" system.

From the results reported in relation to there being a difference between male and female values, the following conclusions are drawn:

1. The locus of control for males in all groupings was more internal than that of females on the Rotter I-E. While only the total male and female groupings were significantly different, the trend for males to be more internal was possibly due to the military occupation being a traditional male job, designed for and by males.

1. A larger study should be done using a variety of value instruments such as the Rokeach, Allport-Vernon-Lindzey, along with the Rotter or Rucker.
2. Since the military is authoritarian, competitive, and demands conformity to standards, it may be informative to examine these characteristics as they relate to success in the military for men and women.
3. Since the study focused on the military as an occupation, and examination of male and female values as they pertain to different jobs within the military may prove fruitful, The requirements of different jobs or Air Force Specialty Codes (AFSC) could be influenced by values.
4. Do females have different support needs as compared with males? If so, how do male or female supervisors differ in their facilitation of these needs and do these measures compare with other measures taken of civilian subjects.
5. Examine the critical adjustment times for men and women when entering the military and how the military helps in the acculturation process.
6. Determine if goal oriented women are more competitive and have fewer adjustment difficulties in the military environment than those who are not as goal oriented or competitive.
7. Determine what environmental factors contribute to and affect the Rotter I-E scores and if these scores can be influenced by the manipulation of these factors leading to possible behavior and attitudinal changes.

IMPLICATIONS

In the span of one decade, the Air Force as a profession has experienced some revolutionary changes with more women entering the service. The numbers of women have increased, uniforms have changed, more career fields are open to women, and the change in the policy regarding pregnancy are just a few of the many changes. These changes have had a positive effect on women joining and staying in the Air Force, but the final decision of how and where women will be made by the American people through their representatives in Congress. Until legislative changes are made, Air Force managers must continue to accomplish the Air Force mission with the human resources available and within the limits of present laws. The present study provides Air Force managers with information which will be useful in determining the effective and efficient ways of utilizing these human resources.

The results of the study suggest there needs to be more extensive research done on values and military personnel. It is clear that the environment of the Air Force has a significant influence on the values of men and women. In spite of the fact that when men and women enter the Air Force, they are as similar in terms of values and philosophies of life and differ only physiologically. But as time passes, they change to become more similar in locus of control of reinforcement.

The ultimate reason for the present study was to provide Air Force managers with information regarding personnel policy-making decisions--it has accomplished

2. Females over time, four years or more in service, became more internal in their philosophies of life. While not substantially significant, this was attributed to acculturation. Females gained knowledge of the system and were better able to influence their future. This change was thought to be influenced as well by what Lefcourt (1973) considered as an individual need to control the situation. Women learn the skills and behavior by observing men and over time become competitive for jobs and promotions.

3. In terms of female recruitment, it was noted that consideration should be given to the external scores of females in early service years. All recruits require adjustment time when entering the military. The length of time to adjust or shift to internal beliefs may affect retention rates. If these adjustment factors could be isolated, adjustment to military life could be facilitated.

4. Essentially, there were few differences in the male and female military members in the study. This suggests that values are not relevant to conflict resolution between males and females in the military, when values are basically the same.

5. One conclusion which was not expected at the beginning of the study was the significant difference between males with one year of service or less and males with four years or more of service.

The study was designed to gain information that would aid Air Force managers in policy making positions when evaluating policies relating to females. However, this assumption was based on the premise that females differed from males in terms of values. The study found that not to be the case, thus the argument is not one of, if people share the similar values, they will live in harmony. Rather, it is that similar values equate to similar needs and when there is limited opportunity to satisfy these needs, i.e., competition, there is conflict.

The Air Force is satisfying its mission requirements and women are an essential part of meeting that mission requirement. Men and women in the Air Force today are similar, in terms of values and philosophy of life as measured by the Rotter I-E, and it seems this was achieved before entering the Air Force. The difference in Rotter I-E scores at entry and later in service may be due to acculturation on the part of the females.

There are greater similarities than differences, but the fact that there is conflict and resistance to females leads to a possible conclusion that the differences are a point of concern. These differences receive more attention from both males and females and have social norms attached to them which are as deeply embedded as other values and cause tension when the outgroup moves into the ingroup's "territory."

RECOMMENDATION

There have been numerous studies of leadership styles, management approaches, and attitudes of personnel in the military environment. However, there is a continuing need to expand this research area to include how these subjects and people interrelate. The following recommendations may be useful:

that goal. While the findings are by no means conclusive, they do require strong consideration prior to the implementation of future personnel policies.

The observations of the study are but the beginning in the quest for a better understanding of values and how they impact on recruitment, retention and performance of duties by all personnel. In essence, the preliminary findings are promising and strongly suggest that psychosocial factors motivating Air Force members require greater consideration in the management of the Air Force's most valuable resources--people.

BIBLIOGRAPHY

1. Allport, Gordon W. 1958. The Nature of Prejudice. Garden City, New York: Doubleday.
2. Allport, Gordon W., Phillip E. Vernon, and Gardner Lindzey, 1951. Manual: A Study of Values. New York: Houghton - Mifflin.
3. Lefcourt, H. M. 1973. "The Function of the Illusions of Control and Freedom", American Psychology, 28:417-425.
4. Phares, Jerry E. 1957. "Expectancy Changes in Skill and Chance Situation", Journal of Abnormal and Social Psychology, 54:339-324.
5. Rokeach, Milton. 1968. Beliefs, Attitudes, and Values. San Francisco: Jossey-Bass, Inc.
6. Rotter, Julian B. 1966. "Generalized Expectancies For Internal Versus External Control of Reinforcement", Psychological Monographs: General and Applies, 1:1-28.
7. Spranger, Edward. 1928. Types of Men. Germany: Max Niemeyer Verlag, Halle (Scale).

Analyzing The Army COHORT Experiment:
Some Critical Questions

Robert P. Kane
Northwestern University
Evanston, Illinois 60021

Abstract

As part of an effort to improve cohesion, the Army is radically changing the way it will man the system, as it moves from an individual to a unit replacement program. This paper proposes alternative models for conceptualizing systems of cohesion. These models suggest possible avenues for the creation of research questions that will help assess the results of the Army's endeavor to enhance cohesion.

The Army today is on the threshold of making a significant change to its manning system. The change involves orienting its system from one that focuses on replacing individual soldiers to one that replaces individual units. The early stage of this changeover has started in a company level replacement program called Project COHORT (Cohesion, Operational Readiness and Training). This program involves twenty-one companies that are projected to remain together for a three year period. The formulation of COHORT units is the first step in the development of an American prototype of the British regimental system, where soldiers will be assigned within the same regiment throughout an entire career. These movements by the Army can be seen as efforts to maximize both the short and long term benefits of stable and cohesive units. Yet the question must be asked: is cohesion solely a product of the organization's ability to align individuals into groups that support organizational goals? This paper puts forth the proposition that the development of cohesion can be viewed as a single process, but with multiple factors influencing the eventual outcome. Two different systems for looking at the development of cohesion will be presented and the significance of implementing each system will be discussed.

INPUT OR THRU-PUT: CONCEPTUALIZING SYSTEMS OF COHESION

A sociological appraisal of social cohesion can be addressed using contrasting methods. In the first, cohesion can be viewed as a dependent variable, with its outcome a function of the social attitudes and backgrounds of individual members of the organization. Seen in this way, the system is open and a function of "input". In the second method, cohesion can be studied as an independent variable, resulting from the operation of the formal organizational process. From this viewpoint, the system is closed or "thru-put". Development of cohesion, then, can be conceptualized along two different dimensions as input and thru-put.

In the input system, social background and attitudes are seen as influencing the outcome of social cohesion.

The social background of group members can be defined as primarily the belief systems of the individuals that make up the group. While social background can encompass a wide array of variables such as ethnicity, sex and social origin, belief systems cut across, and thereby influence all categories. Belief or value systems are important, for they constitute the normative basis for behavior and help define the degree of expected involvement of the individual within an organization. Thus, an input system is concerned with the set of beliefs "carried in" by the individuals as they enter the organization, since they will affect the "outcome" of social cohesion.

We can view the second system in a similar manner. If the formal organizational process is seen as influencing social cohesion, then the structure becomes primarily responsible for its product and the issue is one of thru-put. The structure can be defined as the organization, which trains and socializes new members to appropriate levels, independent of their social background and level of involvement of group members. Goffman (1961) defined this structure as a "total institution", whose formally administered life-style made it possible to change individual members. In this system, then, the individual's social background is considered less important and the system itself becomes the critical factor.

The Importance of Cohesion

As previously noted, both systems seek a single outcome, enhanced cohesion. This outcome is grounded in the maxim that combat efficiency is increased by strengthening the bonding between group members. An understanding of the linkage between cohesion and increased unit effectiveness will help explain why the Army seeks to develop unit cohesion.

Cohesion is normally defined as the degree to which members of a group are attracted to other group members. This means many things to many groups of researchers. Those in organizational behavior suggest this attractiveness motivates members to contribute to the group's welfare and objectives, and to participate in its activities. (Festinger, 1953). Sociologists posit cohesion can also be seen in normative terms, where the social experience of the group process creates values and norms that support a larger social structure. Durkheim (1965) saw that the "collective effervescence" of the group causes individuals to perform to a greater extent than if acting outside the group. In other words, acting on behalf of the group is viewed as acting in the "realm of the sacred", which motivates members to do more than they would by themselves.

Stouffer and his colleagues (1949) refocused the importance of primary group processes in the combat environment, citing the solidarity a soldier experiences with his fellow comrades. Further studies arising from World War II reiterated this point. The ability of the group to override existing civilian values was seen by Grinker and Spiegel, (1963) in their studies of highly cohesive bomber crews. S.L.A. Marshall (1964)

noted the presence of other comrades and the ability to communicate with them was the single greatest motivator in combat among infantrymen in World War II. Similarly, Shils and Janowitz, (1975) studying the Wehrmacht during this same period, found the small, cohesive unit was a formidable support system that resisted disintegration despite the presence of extraordinary pressures.

Thus, the "collective effervescence" present in cohesive units is important because the Army sees it is a force multiplier. Put another way, a highly cohesive unit develops more combat power than a similar unit that is less cohesive.

THE COHORT UNIT: A THRU-PUT SYSTEM

A unit replacement system, stressing the organization's ability to produce motivated soldiers while negating social background, is essentially a thru-put system. Viewed in this way we can learn much from the experience of other thru-put systems. The national prison system, for example is tasked (as a treatment oriented facility), to produce useful and productive persons that can be returned to society fully rehabilitated. (Berliner, 1980). Here the organization must disregard social background and assume it can align the norms and values of any prisoner to match those required by both the the organization and society. Although definitive results on prisons are still required, careful comparative studies in this area would be most beneficial in gaining a better understanding of possible outcomes of the Army's thru-put system.

Essentially, the thru-put system is the "organizational reality" experienced by the individual soldier working in some immediate social situation. Janowitz (1974) suggests this "reality" for the soldier is composed of a combination of four major factors: the nature and type of his weapon system, the organization and type of replacement system, the nature of the threat he must face, and finally, the performance of his military leaders. These factors are the framework that point to the major presumptive advantages the Army expects to gain by implementing the COHORT system.

The first presumptive advantage is the COHORT unit's ability to adapt more quickly and proficiently to new and complex weapon systems. As the Army moves to these systems, the tasks required for efficient operation of the system involve greater complexity and increased integrative behavior. These fighting tasks may take longer to master and require extensive team skill refinement. Today's tank crew, for example, must proceed through a manifold set of qualifications to be declared proficient. Many new weapon systems require this same integrative behavior among group members. In effect, new technology means group proficiency has taken on greater significance than individual proficiency. Cohesive units, because of a common understanding of member abilities, will more quickly overcome individual efficiencies that will aid in mastering intricate group tasks.

The second advantage in this area is important, since it can be assumed that a unit replacement system directly reinforces the first factor. A COHORT system will allow teams that have attained proficiency together to remain as units for longer periods of time. Once complex tasks are learned, they will not have to be relearned because of personnel turnover. The combat proficiency of the unit is maintained for a longer period of time than with the individual replacement system.

The third presumptive advantage is realized in a deterrent situation. Because of the proximity of danger, and the visualization of the enemy being difficult to conceptualize in a peacetime situation, group reinforcement may keep alive the "unseen enemy". Also, as the nature of threat becomes more real, the need for cohesion increases to withstand the actual or perceived stress of battle.

The fourth implied advantage of cohesive units is that they reduce the social distance between the leader and his men, thereby sensitizing leaders to the needs and expectations of the unit. Limited interaction with subordinates is characteristic in a "segmented" military. This tendency is reduced in the COHORT unit, because the interaction is continuous and involves both formal and informal situations. Therefore, a normative commitment develops for the leader to share the concerns and hardships of his men, together with a reciprocal commitment for them to support him. Thus, the COHORT unit gives more flexibility to leaders allowing them to become more proficient at their jobs. Because they have more time with subordinates, leaders will be perceived as doing more than "just going through the motions".

In addition to these factors, two others seem relevant in assessing the COHORT experiment. First, cohesive units can develop and enhance unit pride through gaining an understanding and appreciation of the unit's history. A sensitivity towards unit heritage is important, because it intertwines individual and unit reputations. If an heroic tradition is established, soldiers can more easily define their relationship towards the unit through ideal role models. Yet, association with a single unit over time does not guarantee a soldier will embrace its traditions. Secondly, the network of interpersonal relationships developed within a COHORT unit will allow the soldier to begin to look at his unit as both a home and a community more than just a place to work. Even married soldiers can be integrated, because their families become members of this community, thereby reducing conflicting demands between the soldier's job and his family. This reduction of competing demands allows an individual's commitment to the military to remain focused.

These expected outcomes can be used to frame research questions that measure the effectiveness of the COHORT experiment. Consideration of these factors indicate the COHORT

system's ability to enhance cohesion. If success is less than optimal, one may want to consider factors outside the sphere of organizational influence.

AN ALTERNATIVE TO THE COHORT APPROACH: THE INPUT SYSTEM

The input system can offer its own unique contribution to the development of cohesion. The input system emphasizes social background, the principle component being the belief system brought in by the individual. Belief systems help define the degree of involvement of the members in the organization. Etzioni (1975) stressed the involvement of individual members, and the organization's power to influence those individuals must consistently match. If this does not take place, the individual will not see the organization's power as legitimate, and the organization loses its ability to influence its members. In this situation, members may form cohesive groups that do not support the goals of the organization, as was the case in Vietnam, where excessive drug use and fragging incidents took place. (Moskos, 1975). Faris (1977) found these groups highly cohesive yet not sustaining organizational goals.

Today, the Army must be aware of the possibility of molding cohesive yet disarticulated groups. Members of American society eligible to enter military service may currently question the role and utility of many traditional institutions. In fact, some of the soldiers entering the military may already be alienated against many authority figures within the institution. (Westbrook, 1980). Thus, this problem may be conceptualized as one of understanding the civic consciousness of the individual recruits. Although this means nothing more than the individuals having a concept of the benefits and responsibilities of being an American citizen, many recruits may emphasize only the benefits and not the responsibilities of citizenship. Thus, some form of civic education may be needed to insure that the proper degree of involvement by the members of the organization is understood and maintained.

Finally the input system's core assumption is that unit cohesion is dependent on the quality of the input into the organization. For the Army, this implies that the recruits must come from a representative cross section of American society reflective of the national social composition. In his studies on the All Volunteer Force, Moskos (1981) suggests that "our concern must be with the chemistry of unit cohesion that requires an optimum blend of talents and backgrounds."

Concluding Remarks

This paper has presented two different dimensions for conceptualizing systems of cohesion. These systems were presented as ideal models, but this is not the case. Rather, these systems form the limits within which organizations operate. Because organizations exist within a continuous social system, there can never be a system that is totally input or thru-put.

The COHORT experiment has been viewed in terms of the thru-put system and its ability to enhance cohesion. There are no reasons to believe that the Army will not benefit to some extent from the development of more cohesive units. Yet, two questions remain unanswered that are central to the issues presented. First, how successful is a thru-put system when it has limited control over the quality of its input? Secondly, under what conditions do cohesive groups approve or support the goals of the organization?

REFERENCES

- Berliner, A. Rehabilitation of the imprisoned offender: myth or reality. Corrections Today, 1980, 42, 78-79.
- Durkheim, E. The Elementary Forms of the Religious Life. New York: The Free Press, 1965, pp. 240-245.
- Etzioni, A. A Comparative Analysis of Complex Organizations. New York: The Free Press, 1975.
- Faris, J. Alternative perspective to Savage and Gabriel. Armed Forces and Society, 1977, 3, 457-492.
- Festinger, L. Group attraction and membership. In D. Carwright and A. Zander (Ed.) Group Dynamics: Research and Theory. Evanston: Row, Peterson, 1953.
- Grinker, Roy R. and John P. Spiegel. Men Under Stress. New York: McGraw-Hill Book Co. Inc., 1963.
- Goffman, E. Asylums: Essays on the Social Situation of Mental Patients and Other Inmates. Chicago: Aldine Publishing Co., 1961.
- Janowitz, M. Sociology and the Military Establishment. Beverly Hills Calif.: Sage Publications, 1974.
- Marshall, S.L.A. Men Against Fire. New York: William Morrow and Co., 1964.
- Moskos, C. Making the All-Volunteer Force Work: A National Service Approach. Paper presented for Woodrow Wilson International Center for Scholars. Washington, D.C., March 1981.
- Moskos, C. The American combat soldier in Vietnam. Journal of Social Issues, 1975, 31, 25-37.
- Shils, E. and M. Janowitz. Cohesion and disintegration in the Wehrmacht in World War II. In E. Shils and M. Janowitz (Ed.) Center and Periphery Essays in Macrosociology. Chicago: The University of Chicago Press, 1975.
- Stouffer, S., Suchman, E., De Vinney, L., Star, S., and Williams, R. The American Soldier (Vol. 1) Adjustment during Army Life. Princeton, N.J.: Princeton University Press, 1949.
- Westbrook, S. Sociopolitical alienation and military efficiency. Armed Forces and Society, 1980, 6, 170-189.

HOSTAGE PSYCHOLOGY: THE STOCKHOLM SYNDROME, TRANSFERENCE,
AND IDENTIFICATION WITH THE AGGRESSOR

Victor M. Solomon, PhD, Lt Col, USAFR

Headquarters, Air Reserve Personnel Center
Office of the Command Chaplain
Denver, Colorado 80280

Clinical Director, The Counseling and Psychotherapy Center,
Fair Lawn, New Jersey 07410

Abstract

The literature on hostages was surveyed to determine the Stockholm syndrome as an effect of the hostage experience reciprocating with the mental states, attitudes and behaviors of hostages during their captivity. Thirty-five former hostages were interviewed to ascertain empirically that the Stockholm syndrome is a direct result of the hostage experience, and that the ancillary phenomena of transference and identification with the aggressor were also affected by the hostage experience. It was also found that hostage attitudes toward their captors underwent changes from negative to positive, and protective behaviors such as rationalizing the captor's actions and defending them against the authorities developed. Religious commitment by the hostage and physical or verbal abuse of the hostage may mitigate the Stockholm syndrome.

Hostage-taking has become a serious and familiar problem with global dimensions in little more than a decade (Clyne, 1973). From the hijacking of jet airliners with passengers held for ransom to the protracted captivity of American hostages in Iran, the emotional impact of a hostage experience has touched the lives of thousands of innocent victims and their families. However, the psychopathology of the hostage experience has attracted little scientific attention, perhaps reflecting the recency and novelty of the condition (Hillman, 1981). The increase of terrorism has been both in frequency and violent intensity. Transnational affiliation among terrorist groups has enabled them to receive better training and funding, as well as unprecedented access to complex and sophisticated new weapons technology. Terrorism and hostage-taking thus pose an increasingly greater threat to the individual, the political and economic processes of society, and the quality of international relations (Legum, 1973; Stechel, 1972; Sterling, 1978). Kobetz predicted in 1975 that "hostage-taking incidents ... will increase in intensity and the degree of serious harm which perpetrators inflict on victims" (1972).

The history of hostage-taking goes back to ancient Greece, Rome and the lands of the Norsemen. Middendorf (1972) sketches the violent story of this, at times, colorful and popular activity through the Middle Ages when the Barbary pirates and their European adversaries vied with each other for first place. Barbarossa and Richard the Lion-Hearted played their parts, and Frederick the Great also took hostages to score a political point. Hostage-taking has been an instrument of unconventional warfare throughout the ages, and in every war. Recently, however, it has assumed more menacing proportions in terms of frequency, intensity of barbaric disregard of human life and inclusivity of victims.

The literature relevant to the hostage phenomenon from a psychological perspective is meagre. However, psychological studies on prisoners of war and concentration camp victims have contributed toward an understanding of the problem (Eitinger, 1961; 1962; Berger, 1977; Barocas, 1975; Nederland, 1968; Andersen, 1975). The stress of captivity potentiates certain psychological reactions and behaviors which are, to a large degree, adaptational in character, in order to reduce stress. White (1974) classifies defense mechanisms in the category of adaptation. The hostage experience, though relatively brief, can be as stressful as lengthier captivity encounters (Klonoff, et al, 1976; Eitinger, 1961; Cole, 1980; Jacobson, 1973; Ochberg, 1978).

The defense mechanisms employed by hostages include denial of reality, counter-phobic mechanism, intellectualization, creative elaboration, and reaction formation/identification with the aggressor (Tinklenberg, 1977). Bettelheim (1943) observed identification with the aggressor among some concentration camp victims who assumed behaviors typical of the Nazi guards. This defense mechanism facilitates anxiety reduction by shifting victim's role from oppressed to oppressor (A. Freud, 1966). Patty Hearst is a patent example of this phenomenon. Moreover, shifting allegiances between captives and captors is a fact of terrorism, and victims may be neutral, if not hostile, toward the authorities.

The multiplex hazards of sustained stress have been documented by Selye (1956). Captivity is one of the most traumatic experiences a person can endure. It depersonalizes the victim, transmuting him into a mere commodity for barter. In time, a warmer relationship marked by dependence and identification develops. Terrorist groups, aware of this phenomenon, try to separate victims from captors, or change guards frequently. However, in spite of growing feelings of affection for their captors, hostages endure many physiological and psychological sequelae (Ochberg, 1978).

Captivity is, at core, the deprivation of freedom of movement (Eitinger, 1977). Human dignity and self-esteem are depreciated. Predictability is undermined and a hostage is, thus, deprived of an essential sense of security.

The Stockholm syndrome and related reactions to captivity are best explained by Freudian theories pertaining to ego

defenses (A. Freud, 1966). When other defense mechanisms such as denial of reality have run their course, the Stockholm syndrome emerges. Three phases were observed in the 1973 Stockholm Kreditbank robbery, when hostages were held for 131 hours: the hostages developed positive feelings toward their captors when time elapsed without marked physical or verbal abuse of the hostages; negative feelings toward the authorities emerged as captive identification with the captors developed; captors appeared to reciprocate positive feelings (Strentz, 1979).

Hacker (1976) suggests that physical propinquity and interdependence between captor and captive promote a co-victim feeling and a sense of emotional belonging. A "we" feeling is generated by perceived common danger; i.e., obstinacy of authorities who refuse to comply with the terrorist's demands may lead to the deaths of both hostages and terrorists. An infantile condition of helplessness leads the captive to surrender and obtain relief. The captor's aggression is contagious, and his characteristics are internalized by the hostage (Hacker, 1976). Parents and other agents of society exercise similar powers in the process of socialization training of children by means of object-oriented and love-oriented discipline manipulation, exploiting the child's dependency and approval-seeking needs (McDavid and Harari, 1968). The captor, it has been shown, does not remain unaffected by the relationship.

A genuine fondness existed between the female hostages and the Stockholm Kreditbank robbers, which continued after their release. In other hostage instances, stewardesses visited their imprisoned erstwhile captors regularly long after the event. The literature cites many similar situations.

Method

The Stockholm syndrome, transference and identification were found to be closely connected with the hostage experience throughout most of the literature surveyed for this study. With few exceptions (Hillman, 1981; Modlin, note 1), writers in the field have viewed the mental states, attitudes and behaviors associated with the Stockholm syndrome as normative. Since the relevant published information on this subject is exclusively theoretical, documentary or clinical, the researcher elected to determine empirically the actual presence of the Stockholm syndrome, transference and identification with the aggressor in the hostage experience. The design conforms to an historical and interview research model with computer analysis as back-up for the general discussion.

Subjects

Thirty-five former hostages, 22 males and 13 females were interviewed. Sixteen were captives of the Hanafi Muslims in the Washington, D.C. 1977 incident, 14 were held by the PFLP in Jordan in 1970, and five were recently-returned hos-

tages from Iran.

Instrument

A 41 item structured questionnaire was constructed by the researcher owing to the unavailability of standard tests on the subject. The Likert-type scale was developed from a pilot study with a subgroup consisting of five randomly selected S^s as part of an effort to control for validity and reliability.

Procedure

All subjects were interviewed by telephone. Data base information was followed by input concerning 1. transference related items, 2. identification issues and 3. Stockholm syndrome information, all of which were culled from the literature. Additional items concerning religious commitment and quality of treatment by the captors were explored. The responses were recorded, and numerically rated from 1 to 5, and weighted as such without criterion keying.

Results

The data produced by the interview were computer scored for statistical crosstabulation by means of the SPSS computer software system. At a .05 level of significance, the chi-square test indicated support for the near-consensus view expressed in the literature that 1. there is a Stockholm syndrome experienced by hostages; 2. that transference and identification with the aggressor are included in that syndrome; 3. that religious commitment, and quality of treatment of the hostages significantly affected the reaction to captivity in terms of the hostage syndrome.

Discussion

The theoretical/clinical observations hypothesizing a Stockholm syndrome and related phenomena were supported by the respondents in the survey, and further strengthened by the results of the statistical computer back-up to the discussion. Attitude change toward captors proceeds from negative to positive with rationalizations to defend them against the authorities, except when the hostages are subjected to physical or verbal abuse, or in the case of hostages with strong religious commitment/practice.

Conclusion

With the proliferation of terrorism and hostage-taking during the past decade and a half, the need for this research assumes critical significance. The increased probability of abduction dictates a greater awareness of the problem and how to deal with it before, during and after captivity. The results of this research represent a small step in that direction.

Reference Note

1. Modlin, H.C., M.D., The Menninger Foundation. Personal Communication.

REFERENCES

- Anderson, R.S. Operation homecoming: Psychological observations of repatriated Vietnam prisoners of war. Psychiatry, 1975, 38, 65-74.
- Barocas, H.A. Children of purgatory: Reflection on the concentration camp survival syndrome. International Journal of Social Psychiatry, 1975, 21, 87-92.
- Berger, D.W. The survivor syndrome: A problem of nosology and treatment. American Journal of Psychotherapy, 1977, 31, 238-251.
- Battleheim, B. Individual and mass behavior in extreme situations. Journal of Abnormal Social Psychology, 1943, 38, 417-452.
- Clyne, P. An anatomy of skyjacking. Scranton, Penna: Intext, 1973.
- Cole, D. Why a hostage cannot forget. Newsweek, May 19, 1980, 17.
- Eitinger, L. Pathology of the concentration camp syndrome. Archives of General Psychiatry, 1961, 5, 371-379.
- Freud, A. The ego and the mechanisms of defense. New York: International Universities Press, 1966.
- Hacker, F.J. Crusaders, criminals, crazies: Terror and terrorism in our time. New York: Norton, 1976.
- Malman, R.G. The psychopathology of being held captive. American Journal of Psychiatry, 1981, 138, 1193-1197.
- Jacobson, S. Individual and group responses to confinement in a sky-jacked plane. American Journal of Orthopsychiatry, 1973, 43, 459-469.
- Klonoff, H., McDougall, G., Clark, C. et al. The neuropsychological, psychiatric, and physical effects of prolonged and severe stress: Thirty years later. Journal of Nervous and Mental Disorders, 163, 246-250.
- Kobetz, R.W. Hostage incidents: New police priority. Police Chief, 1975, 42, 30-37.
- Legum, C. The rise of terrorism. Current, 1973, 147, 3-9.
- McDavid, J.W. and Harari, H. Social psychology: Individuals, groups, societies. New York: Harper and Row, 1968.
- Middendorf, W. Taking hostages and kidnapping. Kriministik, 1972, cited in M.D. Beall. Hostage negotiations. Law Enforcement Journal, 1976, 3, 23-44.
- Niederland, W.G. Clinical observations on the survivor syndrome. International Journal of Psychoanalysis, 1968, 49, 313-315.
- Ochberg, P.M. The victim of terrorism: Psychiatric considerations. Terrorism, 1978, 1, 147-168.
- Selye, H. The stress of life. New York: McGraw-Hill, 1956.
- Stechel, I. Terrorist kidnapping of diplomatic personnel. Cornell International Law Journal, 1972, 5, 189-217.
- Sterling, C. Terrorist Network. Atlantic Monthly, 1978, 242, 37-43, 46-47.
- Strentz, T. Law enforcement policy and ego defenses of the hostage. FBI Law Enforcement Bulletin, 1979, 48, 2-12.
- Tinklenberg, J.R., Murphy, Peggy and Murphy, Patricia. Adaptive behavior of victims of terrorism. in R.D. Crelinstein (ed.). Dimensions of victimization in the context of terroristic acts: Final report. International Centre for Comparative Criminology. Montreal: University of Montreal, 1977, 93-107.
- White, R.W. Strategies of adaptation: An attempt at systematic description. in G. Coelhod, D. Hamburg and J. Adams (eds.). Coping and adaptation. New York: Basic Books, 1974, 47-68.

POSTLUDE: The author of this presentation is completing a major research project: HOSTAGE PSYCHOLOGY: CAPTIVE, CAPTOR AND CAPTIVITY, for publication. Detailed statistical treatment of the research, with cognate statistical data and graphica are available.

A Processing Resource Explanation of Subjective Dimensions of Operator Workload

William L. Derrick
United States Air Force Academy

ABSTRACT

A multiple resource model of processing resources (Wickens, 1980) guided construction of four modal tasks that demanded different configurations of resources. Twenty-two subjects performed these tasks in all single and dual task combinations, then rendered paired comparison judgments of workload similarity. Multidimensional scaling analysis of these judgments produced three dimensions that seemed to underlie subjective perceptions of workload. These dimensions were explained by 1) task performance decrements predicted by the model, 2) task structure differences, and 3) perceived stress.

INTRODUCTION

Background

Moray (1982) has recently focused attention on the association between classes of task characteristics and the subjective experience of operator workload. Categorizing task into physical, cognitive, manual-control, and time-stress areas, he identifies characteristics such as a tracker's requirement to generate lead, the number of alternative solutions in a decision making task, the scarcity of time, and so forth as reliably producing the perception of load. Moray concluded the article by noting that any proposed model of mental workload that attempts to explain operator reaction to these and other characteristics will be both complex and multivariate.

Although exploring the relationship between task characteristics and mental workload can provide valuable insights, it leads to several problems. First, an enormous number of task characteristics can currently be identified, and as technology gives us new equipment and tasks, they can only proliferate. Second, a combination of task characteristics may produce subjective experiences that are only weakly related to the experiences associated with the characteristics in isolation. How many combinations can reasonably be examined? Finally, examples are reported in the literature (e.g., Herron, 1980) in which task characteristics were modified and subjective experiences changed concomitantly, but system performance either did not change or changed in an unpredictable way. This can be rather disconcerting for a system designer attempting to use subjective data in the design of an effective system.

An alternative to this approach starts with a performance-based theory of workload and asks how do perceptions of mental workload relate to explanations of performance changes. This performance-based theory was proposed by Wickens (1980) and deals with the construct of attentional resources. Although Wickens states that task performance requires these finite resources and that decrements in performance occur as the resources are expended, he contends that many separate entities of attentional resources exist. Specifically, these resources are defined by stages of processing, modalities of input and output, and codes of

central processing (Wickens, 1980). When tasks that demand approximately the same constellation of resources are time-shared (e.g., they compete for resources), performance decrements will occur. Alternatively, time-shared tasks that utilize separate constellations of resources will be performed effectively.

This model provided the hypothesis to be tested in this study. Specifically, the underlying structure of workload perceptions can be explained by resource demand when resources are defined by stages, modalities, and codes.

Approach

To study the relationship between the subjective and performance-based measures of workload, multidimensional scaling (MDS) methodology was employed. Four tasks were created, the successful performance of which demanded resources from different parts of the Wickens model. Two versions of each task, one easy and one hard, were constructed. The four easy tasks were then combined in all possible pairs to form ten dual task combinations, some of which competed for resources while others did not. These 18 tasks (4 easy, 4 hard, 10 dual tasks) were first performed by subjects and then served as stimuli for paired comparison judgments of workload similarity. MDS analysis of these judgments produced the attributes or dimensions underlying perceptions of workload. Location of tasks in this MDS space indicated how differential resource cost drove opinions of workload.

METHOD

Subjects

Twenty-two right handed males from the University of Illinois, aged 18 to 22, served as subjects. All reported no experience with the types of tasks used in this study. Each was paid \$3.50 per hour plus performance bonuses for his participation.

Tasks

Tracking (T). A one-dimensional critical tracking task required subjects to keep a moving error cursor centered on a stationary horizontal bar. The cursor, which traveled in a vertical plane, required fore-aft movements of a spring-centered joystick to null the error. The parameter lambda was set at 0.95 for the easy version and 1.90 for the difficult version. Control was with the left hand, except when time-shared with itself, in which case both hands were used. According to the multiple resource model, this task should demand resources dedicated to spatial codes, visual input modalities and manual response modalities, and to stage-related resources associated with response execution.

Visual Search (V). Subjects searched for the target word NOW that was embedded in a 4 x 8 array of upper case letters and nonletters. Target probability was set at 35%. The array contained 15% nonletters for the easy task and all letters for the difficult task. Detection required squeezing a trigger on a joystick with the right hand. When time-shared with itself, subjects searched for two targets, NOW and ONE, with ONE requiring a left hand response. The composite resource demands of this task should be defined by verbal codes, visual input modalities, manual response modalities, and perceptual/central resources associated with stages of processing.

Auditory Sternberg (S). Subjects retained in memory either two (easy task) or five (hard task) memory set letters for each trial. Subsequent auditory presentations of letters to the right ear required a vocal YES if the heard letter came from the memory set, a vocal NO otherwise. Memory set letters occurred 50% of the time. When time-shared with itself, a memory set of two letters was presented to each ear with YES the appropriate response only when letter and ear were correctly associated. This task should demand resources dedicated to verbal codes, auditory input modalities, vocal response modalities, and resources primarily from the perceptual/central structure.

Tone Judgment (J). A 1200 millisecond tone burst to the left ear required subjects to position an arrow on a vertical scale associated with that tone frequency. The easy task required absolute judgment of 4 tones (400 Hz, 480 Hz, 650 Hz, and 850 Hz) while the difficult task added a tone at each end of the scale (325 Hz and 1000 Hz). Left hand joystick manipulations of zero order dynamics positioned the arrow. Tones to both ears, two scales, and two joysticks were employed when the task was time-shared with itself. Performance on this task should demand resources associated with spatial codes, auditory input modalities, and manual response modalities. Unlike the other tasks, the judgment and control aspects of this task should require both perceptual/central and response dedicated resources.

Design and Procedure

All subjects performed each of the 18 tasks during three sessions, two practice and one experimental. Each task trial lasted two minutes. During task trials, bonus payments based upon the previous day's performance were used to encourage equal concern for both tasks. For the experimental session, task order was permuted to ensure that each task occurred at each of 18 possible serial positions. Subjects received performance feedback after each trial.

Immediately following the experimental session, task ratings were collected. Subjects rated the similarity of task difficulty (assumed to be equivalent to workload) for all possible pairs of the 18 tasks ($n = 153$). Each task was then rated on five unidimensional scales (Input Complexity, Response Complexity, Effort, Feedback Adequacy, and Time Demands) and all were rank ordered according to task difficulty.

RESULTS

Resource Competition

To validate the differential resource demands of these tasks, the 10 dual tasks from Session 3 were analyzed in terms of performance decrements. Performance in the dual task trials was scored as a change in the relevant performance measures using the easy single tasks as a baseline.

For the purposes of this paper, it is sufficient to note that time-sharing performance decrements predicted by the model were obtained. MANOVAs revealed that the decrements were reliably different, and the correlation between rank orderings of predicted and obtained decrements was substantial (Spearman $\rho = .72$, $p < .05$).

Multidimensional Scaling

The paired comparison judgments of task difficulty were analyzed by the Individual Differences Scaling (INDSCAL) program (Carroll & Chang, 1970). This program permitted direct interpretation of the produced dimensions without additional rotation (Kruskal & Wish, 1978). Using the criterion of variance accounted for, a three dimensional solution was selected (55% VAF).

Location of the 18 tasks in the MDS space indicated that Dimension 1 reflected perceptions of task difficulty and workload associated with resource demand. The single task "hard" Visual Search and Autirotty Sternberg tasks were exceptions, primarily because task parameter manipulations did not appreciably increase task difficulty level. Task location on Dimension 2 indicated that workload perceptions were influenced by the input modality required. Highly positive Dimension 2 weights were associated with primarily visual input, slightly positive and negative weights occur with mixed input tasks, and highly negative weights were associated with primarily auditory input tasks.

Correlations between the scaling dimension weights, the unidimensional ratings, and the performance decrements (Table 1) aid in dimension interpretation. The Effort Demand, Difficulty Ranking, and Performance correlations with Dimension 1 indicate that resource demand both influences performance and contributes significantly to subjective workload perceptions. The Feedback Adequacy and Input Complexity correlations with Dimension 2 indicate that the stimuli in the visual tasks were perceived as less complex and providing more inherent feedback than the auditory tasks.

TABLE 1

Correlations of INDSCAL Dimension Weights with Unidimensional Scale Means, and Average Performance Decrements (10 dual plus 4 hard single tasks)

Variable	<u>Dimension</u>		
	1	2	3
Unidimensional Scales:			
Input Complexity	-.17	-.62	-.28
Response Complexity	-.47	-.22	-.56
Effort Demands	-.67	-.32	-.61
Feedback Adequacy	.17	.92	-.38
Time Demands	-.07	.36	-.72
Difficulty Ranking	-.70	-.54	-.55
Performance Decrement	-.80	.07	-.23

Note. The following critical values can be used to evaluate the correlations reported in the table: $r(12) = .661, p < .01$; $r(12) = .532, p < .05$.

The interpretation of Dimension 3, not readily apparent from task location in the scaling space, is suggested by the correlational data. Reliable correlations with perceived Effort, Time Demands, and Response Complexity, but no correlation with Performance suggest subjects are reporting task stress, a commonly reported attribute of subjective workload (Sheridan, 1980).

DISCUSSION

Mapping tasks of known resource demand onto the subjective dimensions of workload indicated that the multiple resource model is highly associated with two of those dimensions. This approach to explaining how perceptions of workload evolve appears to be more parsimonious than the identification and investigation of numerous task characteristics. Utilizing a theoretical framework of relatively few variables, it both ties the subjective experience of load to performance and indicates how these two measures dissociate. Indeed, this use of multidimensional scaling permits one to investigate that very dissociation, and by using the criterion of variance accounted for, indicates for any set of tasks how many unique dimensions of mental workload may exist.

REFERENCES

- Carroll, J. & Chang, J. Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition. Psychometrika, 1970, 35, 283-319.
- Herron, S. A case for early objective evaluation of candidate display formats. Proceedings of the 24th Annual Meeting of the Human Factors Society, Los Angeles, California, October, 1980, 13-16.
- Kruskal, J. & Wish, M. Multidimensional scaling. Beverly Hills, California: Sage Publications, 1978.
- Sheridan, T. Mental workload - What is it? Why bother with it? Human Factors Society Bulletin, 1980, 23, 1-2.
- Wickens, C. The structure of attentional resources. In R. Nickerson & R. Pew (Eds.), Attention and Performance VIII. Englewood Cliffs, New Jersey: Erlbaum, 1980.

INDIVIDUAL DEFINITIONS OF THE TERM "WORKLOAD"

Sandra G. Hart
NASA/Ames Research Center
Moffett Field, CA

Mary E. Childress Jan R. Hauser
San Jose State University
San Jose, CA

ABSTRACT

A study was conducted in which four groups of raters (51 researchers, 28 college students, 12 general aviation pilots, and 26 high school students) assigned 19 possible components of workload to one of three categories: (1) not related to workload; (2) related to, but not a primary component of workload; and (3) a primary element of workload. These ratings were factored to determine the relationships among the items. The analysis yielded seven factors: fatigue/stress, task difficulty, effort, performance/motivation, task type, interest in task, and purpose of task. The 117 participants were clustered on the within-subject standardized factor scores. This analysis yielded seven patterns of responses about the relative primacy of the different factors to different individuals' definitions of workload. The results indicate that patterns of estimating the primacy of components in subjective workload evaluation exist which cross working group lines.

INTRODUCTION

The term "workload", rather than referring to a well-defined, unique and generally agreed upon phenomenon, serves as a convenient label for a number of events, ideas, states, dimensions, and other constructs that are ill-defined and difficult to measure. Although the term is widely used, and there is some agreement on broad definitions, there is considerably less agreement on specifics. When one individual reports working very hard, he or she may be referring to physical fatigue, whereas another person, expressing an equal level of workload, may be referring to the fact that he or she had to do more than expected, even though, on an objective level, little was done. To some, workload is completely defined by the demands of a task, while to others it is defined by the physical and mental effort exerted to meet such demands. The goal of the assessor in measuring workload may vary as well: increments in workload may be relevant to one assessor only if the result is a decrement in operator or system performance, but to another assessor, the physical, emotional, or mental cost to the operator may be the primary concern.

Some dimensions of workload may be independent, whereas others interact with and influence each other. Some components are independent of the operator (e.g., task demands or the environment) whereas others relate primarily to the operator (e.g., motivation, fatigue, or effort expended). Task-related and operator-related factors may either derive specifically from the task at hand or act as covariates or moderators. Finally, there are factors that may be thought of as the result of workload, such as performance. Performance may be the intersection between task demands, effort expended, skills available, and the circumstances under which a task is performed rather than workload per se.

One concept that appears to reflect all of the relevant components is the individual's subjective experience. If a person feels loaded, then he or she is loaded. This feeling may be the only aspect of the constellation of elements that have been considered relevant to workload that is purely

"workload". The subjective experience is, of course, based on many of the other factors, but different elements may be more or less important to one person than to another. For example, time pressure or the number of tasks waiting to be done may be important to one person, whereas successful performance or personal well being may be crucial to another. Thus, the subjective experience of workload may be the result of a weighting process that is based on fairly consistent directions of focus within an individual but on different factors between individuals. The weight or importance that each person places on different factors may be thought of as a filter through which all that occurs and is experienced is transformed. The output of this process is a measurable response that serves as a crude summary or index of the workload that the person experienced. Thus, the impression that one receives about the workload imposed by a situation is determined not only by the type of measure chosen but also by the characteristics of the person from whom the measure is obtained. Measures that are appropriate for evaluating the facets of workload that are crucial to one individual's experience of workload may not reflect another's. A measure of workload is simply a probe which may be applied precisely and wisely or miss the mark entirely either because it is inappropriate for the situation, does not address the questions being asked, or because it does not focus on the aspects of workload that are important to the individual being studied.

If workload is in fact dependent on the inner experience of an operator, then we would predict that an operator's subjective feeling of the relative level of workload would be reflected in objective measures such as: changes in heart rate, increased errors, or inappropriate selection of strategies, as well as in subjective evaluations. The purpose of this study was to determine what factors are important to different people to provide a basis for understanding the impact of differences in subjective experiences on workload assessment. Specifically, we wished to find out: (1) whether individuals agree about the importance or relevance of a number of possible components of workload; (2) whether a limited number of dimensions may describe and predict ratings on groups of components; and (3) whether there are patterns of responses in the importance placed on different factors that are either characteristic of, or cut across, occupational groups.

METHOD

Subjects

This study was conducted with four groups of raters drawn from somewhat different elements of the population. Each group was convened for a purpose other than the current study; thus the size and composition was not experimentally controlled. Group 1 was composed of 51 participants in the Man-Vehicle Systems Research Division's research seminar held at NASA-Ames Research Center. There were approximately the same number of men and women and their ages ranged from 21 to 70 years. The majority of the participants held advanced degrees in psychology and engineering, were engaged in human factors research in aviation, and nearly half were pilots. Group 2 was composed of 28 undergraduate and graduate students in a management psychology class at San Jose State University. Eleven women and 16 men, from 20 to 50 years in age, volunteered to participate. Only two were pilots. Group 3 was composed of 12 general aviation pilots with instrument ratings who also served as paid participants in other experiments. There were two women and 10 men who ranged in age from 22 to 45 years. Group 4 included 26 high school seniors who were selected to participate in the NASA Space Biology Program. They were 17 and 18 years old.

Table 1
Item Descriptions

Item	Title	Explanation
1	Task Demands-Type	Does it matter whether the task demands are primarily physical, cognitive, or perceptual?
2	Task Demands-Amount	Is workload related to how demanding/taxing/hard the task is?
3	Task Complexity	Is the number of different tasks or task elements that have to be attended to relevant to workload?
4	Time Pressure	How important is the amount of time available to perform a task to an estimate of workload?
5	Feedback	Is the presence or absence (or adequacy) of information about how well you are doing during a task significant to workload?
6	Purpose	Is the purpose of the activity (e.g., whether it is "work" or fun) a factor in workload?
7	Task Length	Does the length of time it takes to do a task affect workload?
8	Environment	Is the environment in which a task is performed relevant?
9	Performance	How important is your success in meeting task demands (how well you did or felt you did) in workload?
10	Boredom	If a task is boring (or interesting) is the workload affected?
11	Training	Is the amount of training, prior experience, or familiarity with the task related to workload?
12	Mental Busyness	Is mental busyness--such as short-term mental rehearsal to remember things--relevant?
13	Physical Busyness	Is physical busyness per se (as opposed to the amount or intensity of physical effort) related to workload?
14	Motivation	Is the importance or relevance of a task a significant influence on the workload?
15	Well Being	Does your current state of physical and emotional well-being affect the workload experience?
16	Effort--Type	Does the <u>type</u> of effort (e.g., perceptual, physical, or mental) expended affect workload?
17	Effort--Amount	Does the <u>amount</u> of effort expended affect workload?
18	Fatigue	Is fatigue/sleepiness/tiredness related to the workload experienced during a task?
19	Emotional Stress	Do feelings of anxiety/tenseness/calmness/tranquility/fear etc. contribute to workload?

Test Materials

Each participant was given an envelope with three compartments and a set of 19 cards that presented the 19 components of workload to be evaluated. The three compartments represented three possible levels of relevance to the rater's concept of workload: not related, somewhat related, or a primary factor. Each card was labeled with a descriptor and an operational definition (Table 1).

Procedure

The responses were obtained from each group following a briefing in which the theory and practical application of workload assessment research were presented. The participants were asked to read each card and to respond to the definitions rather than to the one- or two-word label. Each item was to be placed in the compartment that represented the degree to which the rater would take that factor into account when rating or experiencing workload.

RESULTS AND DISCUSSION

Numerical values of 0, 1, and 2 respectively were assigned arbitrarily to ratings of "not related", "somewhat related", and a "primary element" to allow tallying and analysis. Even though we recognized that the data were categorical in nature, we considered that the information obtained from the following statistical analyses warranted their use.

Table 2 Percent of raters who placed each of 19 workload components into each of three relevancy categories. (n=117)

Component	Relevancy categories		
	Not related to workload	Somewhat related to workload	Primary element of workload
1. Task type	17%	45%	38%
2. Task difficulty	06	21	73
3. Number of tasks	06	29	65
4. Time pressure	04	24	72
5. Feedback	16	57	27
6. Purpose of task	20	49	31
7. Task length	15	48	37
8. Environment	13	61	26
9. Performance	08	45	47
10. Interest in task	13	58	29
11. Training	08	43	49
12. Mental busyness	09	41	50
13. Physical busyness	23	46	31
14. Motivation	08	49	43
15. Feeling of well-being	06	51	43
16. Type of effort	13	46	41
17. Amount of effort	08	31	61
18. Fatigue	03	41	56
19. Stress	02	34	64

The most striking result observed in the raw data was the lack of agreement within each original group and across all raters about just what does and what does not influence individuals' perceptions of workload. A summary of the responses may be seen in Table 2. The range of differences and the patterns of responses within each group were consistent across the four groups, although there were a few specific areas of disagreement. Kendall's coefficient of concordance ($w = .077$) indicated no overall agreement among the 117 raters in the way that they responded to the 19 items.

Each of the items was selected as being somewhat or primarily related to workload by at least 75% of the group, providing strong support for the assertion that workload is multi-dimensional. Task-related components, such as task difficulty, complexity, and time pressure were selected as primary elements by almost all of the raters. Person-related components, such as effort, emotional stress, fatigue, and feeling of well being were selected by a large number of raters as primary as well; however there was less agreement about the primacy of these elements. Other factors, such as performance, purpose, and environment, were more frequently selected as related rather than primary elements of workload.

A principle components analysis was performed on the standardized ratings for all 117 participants to determine whether there was a more limited set of dimensions that would account for the responses. The analysis yielded seven factors, selected because they provided the highest loadings and the most interpretable factors (Table 3). These components will be referred to by the following labels: (1) fatigue/stress; (2) task difficulty; (3) effort; (4) performance/motivation; (5) task type; (6) interest in task; and (7) purpose.

Table 3 Principle components analysis of nineteen workload components (Factor loadings)

Components	Factors						
	I	II	III	IV	V	VI	VII
Fatigue	.72	.19	.05	-.07	-.03	.16	-.02
Well-Being	.71	-.13	.03	0	.19	.31	.03
Emot. Stress	.71	.18	.03	.22	-.16	.02	.04
Training	.55	.06	.04	.14	.43	-.12	.11
No. of tasks	.05	.81	.02	.02	.21	-.10	.05
Time pressure	.10	.74	.05	.15	-.13	.05	.11
Difficulty	.21	.62	.32	.06	.07	-.03	-.23
Phys. Busy.	-.06	.27	.78	-.07	-.03	.22	-.04
Amt. of effort	.02	.02	.76	.12	.20	-.18	.33
Mental Busy.	.24	.06	.65	-.06	.19	-.01	-.36
Feedback	.09	0	.04	.76	.08	.28	-.19
Performance	.15	.06	-.05	.76	.04	-.11	.15
Motivation	-.15	.13	0	.58	.08	.25	.46
Effort Type	.07	.01	.20	0	.77	.06	.16
Task Type	-.03	.16	.03	.13	.74	.13	-.27
Interest	.12	-.05	-.04	.12	-.05	.74	.24
Environment	.32	.01	.05	.13	.18	.58	-.11
Purpose	.12	.04	.01	.05	-.02	.11	.80
Task Length	-.03	.45	.21	-.09	.24	.39	.10

A cluster analysis was performed on the within-subject standardized ratings for the seven factors to determine whether there were identifiable patterns of responses about what factors were important to different individuals' definitions of workload. The analysis was performed using an iterative partitioning method based on pairwise score differences (refs. 1-7). The goal was to minimize the variance within each resulting cluster so that similar patterns of responses were grouped together and the resulting groups were maximally different. This analysis yielded seven characteristic patterns of responses. (figure 1) These seven clusters tended to cross socially and occupationally defined group lines; however some groups were better represented by one pattern than another. For example, there were a greater than chance number of college students in Cluster F, high school students in Cluster B, and pilots in Cluster E. There were fewer than chance researchers in Cluster G and college students in Cluster A. The clusters were essentially equivalent in size, as indicated by a nonsignificant difference in the proportion of the total group in any of the seven clusters ($\chi^2 = 2.86$, $df=3$, n.s.).

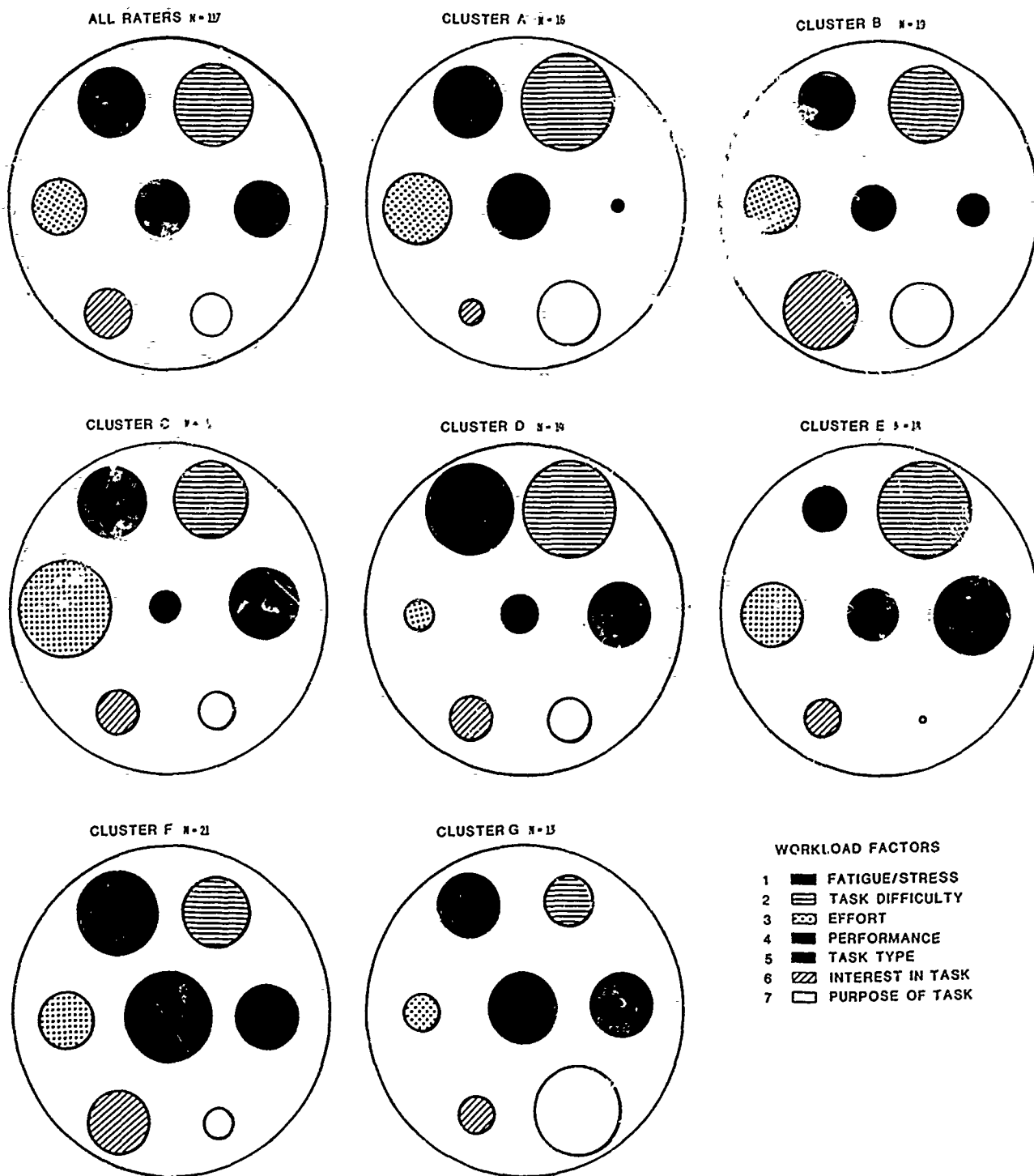
In order to determine whether a person's ratings could be better predicted by knowing from which population he was drawn (e.g., researchers, college students, high school students, or pilots), or by knowing the response pattern cluster to which he or she had been assigned, (e. g., Clusters A-G), a discriminant analysis (ref. 8) was performed on the ratings for the seven workload factors. For 73% of the raters, the importance placed on each of the seven factors could be correctly predicted by knowledge of original group membership, whereas 99% of the raters' responses on the seven factors could be classified correctly according to cluster membership.

CONCLUSIONS

This study clearly demonstrated that people disagree about which of the many possible elements of workload are actually relevant to their own definition and experiences. Some of the components included in the study were selected consistently as primary aspects of workload (e.g., task demands, fatigue, effort, and time pressure), whereas others were selected by some respondents and not by others (e.g., performance and purpose of task). The original 19 elements were found to be characterized by seven principal components. Even this reduction in the number of factors believed to be closely related to workload still leaves an extremely broad range of elements that need to be considered when assessing operator workload. Furthermore, the ratings obtained for the original 19 components, or for the seven principle components, varied considerably among the participants. Seven characteristic patterns of response were found for the factors that included the major differences in importance placed on the different factors by the participants. These seven clusters crossed social and occupational group lines, indicating that information of a general, rather than occupation-specific, nature was obtained. With additional research, it may become possible to identify in advance of research the importance that an operator will place on the different aspects of the situation in experiencing workload. With this knowledge, (e.g., knowing in advance whether an individual will focus on task demands or on physical state), more precise and accurate measurements may be made of the workload imposed by different tasks. This study was exploratory and must be followed by behavioral research to determine whether the factors that people believe they consider when assessing workload are in fact related to their estimates of and reactions to the workload imposed by a variety of circumstances.

Figure 1: PATTERNS OF RATINGS FOR THE 7 FACTORS IDENTIFIED AS WORKLOAD COMPONENTS

(THE DIAMETER OF EACH CIRCLE IS PROPORTIONAL TO RATINGS GIVEN TO COMPONENTS WITH FACTOR LOADINGS GREATER THAN .5)



REFERENCES

1. Blashfield, R. K. & Aldenderfer, M. S. Computer programs for performing iterative partitioning cluster analysis. Applied Psychological Measurement, 1978, 2, 533-541.
2. Childress, M. E. Statistics for evaluating classifications: A new view. Paper presented at the meeting of the Classification Society (North American Branch). Toronto: June, 1981
3. Cronbach, L. J. & Gleser, G. C. Assessing similarity between profiles. Psychological Bulletin, 1953, 50, 456-473.
4. Milligan, G. W. An examination of six types of error perturbation on fifteen clustering algorithms. Psychometrika, 1980, 45, 325-342.
5. Osgood, C. E. & Succi, G. J. A measure of relation determined by mean differences and profile information. Psychological Bulletin, 1952, 49, 251-262.
6. Schoenfeldt, L. F. Utilization of manpower: Development and evaluation of an assessment-classification model for matching individuals. Journal of Applied Psychology, 1974, 59, 583-595.
7. Wishart, D. Clustan user manual (3rd Ed.). Edinburgh, Scotland: Program Library Unit, Edinburg University, January 1978. (Inter-University/Research Councils Series Report No. 47).
8. Rogers, G. & Linden, J. D. Use of multiple discriminant function analysis in the evaluation of three multivariant grouping techniques. Educational and Psychological Measurement, 1973, 32, 787-802.

THE ESSENCE OF PILOT FLYING SKILL

Edward E. Eddowes
Chief of Naval Air Training Staff
NAS Corpus Christi, TX

Basic characteristics of aircraft operations will be examined in this paper to develop a description of the essence of the pilot's flying skill which transcends the explanatory limits of an older notion of pilot skill as hand-eye coordination. The objective of this investigation is to generate an improved appreciation of pilot task requirements and to provoke a reexamination of the goals of flying training programs and of their form and substance. Before proceeding with this analysis of pilot skill, the implications of controlling an aircraft's flight path in three-dimensional space will be considered, and the implications of the requirement for developing plans to accommodate a variety of conditions which may arise during a flight will be described to establish a starting point for the discussion which follows.

CHARACTERISTICS OF FLIGHT OPERATIONS

Aircraft operate in an elastic medium and are controlled by means of a pilot's manipulation of the configuration of the aircraft's control surfaces. The speeds at which aircraft normally operate prevent abrupt changes in flight path and require that such changes be accomplished gradually and smoothly enough to avoid departures from controlled flight. Thus, the pilot must learn the performance limits of an aircraft and become familiar with its maneuvering characteristics. This will enable the pilot to control the aircraft's flight path as required to accomplish all maneuvers within its performance envelope.

In addition to mastering flight maneuvers, a pilot must know the requirements for controlling an aircraft throughout a mission or must have such requirements information available to properly control the aircraft's flight path and manage its operation. Because flying operations often are conducted under variable environmental conditions, it is a practical necessity that a pilot know of a number of alternative means which can be used to achieve mission objectives if for any reason prior plans or requirements information no longer apply due to environmental or aircraft status changes. Consequently, pilots are trained, qualified and periodically evaluated on the discipline, accuracy, efficiency, completeness and safety of the aircraft control and system management procedures they use in accomplishing a missions.

THE PILOT'S DYNAMIC ENVIRONMENT

As a result of operating an aircraft in a dynamic, three-dimensional environment and practicing flying maneuvers, a pilot learns and knows the maneuvering characteristics of the aircraft and, given sufficient information and understanding of the requirements of a mission, can plan (formally or informally) what must be done to satisfy the requirement (Langewieshe, 1944). Through familiarity with an aircraft and knowledge of mission requirements, a pilot can satisfy them accurately, efficiently and safely. To deal effectively with momentary aircraft control tasks, a pilot must project the current status of the aircraft into the near future time-space, evaluate this projection in terms of the remaining mission requirements and take appropriate control actions to continue satisfying them. This kind of behavioral process has been described by Hull (1930) and may be assumed to support the pilot's projection of current aircraft states into the future in order to determine the aircraft control action requirements.

THE ESSENCE OF PILOT FLYING SKILL

Edward E. Eddowes
Chief of Naval Air Training Staff
NAS Corpus Christi, TX

Basic characteristics of aircraft operations will be examined in this paper to develop a description of the essence of the pilot's flying skill which transcends the explanatory limits of an older notion of pilot skill as hand-eye coordination. The objective of this investigation is to generate an improved appreciation of pilot task requirements and to provoke a reexamination of the goals of flying training programs and of their form and substance. Before proceeding with this analysis of pilot skill, the implications of controlling an aircraft's flight path in three-dimensional space will be considered, and the implications of the requirement for developing plans to accommodate a variety of conditions which may arise during a flight will be described to establish a starting point for the discussion which follows.

CHARACTERISTICS OF FLIGHT OPERATIONS

Aircraft operate in an elastic medium and are controlled by means of a pilot's manipulation of the configuration of the aircraft's control surfaces. The speeds at which aircraft normally operate prevent abrupt changes in flight path and require that such changes be accomplished gradually and smoothly enough to avoid departures from controlled flight. Thus, the pilot must learn the performance limits of an aircraft and become familiar with its maneuvering characteristics. This will enable the pilot to control the aircraft's flight path as required to accomplish all maneuvers within its performance envelope.

In addition to mastering flight maneuvers, a pilot must know the requirements for controlling an aircraft throughout a mission or must have such requirements information available to properly control the aircraft's flight path and manage its operation. Because flying operations often are conducted under variable environmental conditions, it is a practical necessity that a pilot know of a number of alternative means which can be used to achieve mission objectives if for any reason prior plans or requirements information no longer apply due to environmental or aircraft status changes. Consequently, pilots are trained, qualified and periodically evaluated on the discipline, accuracy, efficiency, completeness and safety of the aircraft control and system management procedures they use in accomplishing a missions.

THE PILOT'S DYNAMIC ENVIRONMENT

As a result of operating an aircraft in a dynamic, three-dimensional environment and practicing flying maneuvers, a pilot learns and knows the maneuvering characteristics of the aircraft and, given sufficient information and understanding of the requirements of a mission, can plan (formally or informally) what must be done to satisfy the requirement (Langewieshe, 1944). Through familiarity with an aircraft and knowledge of mission requirements, a pilot can satisfy them accurately, efficiently and safely. To deal effectively with momentary aircraft control tasks, a pilot must project the current status of the aircraft into the near future time-space, evaluate this projection in terms of the remaining mission requirements and take appropriate control actions to continue satisfying them. This kind of behavioral process has been described by Hull (1930) and may be assumed to support the pilot's projection of current aircraft states into the future in order to determine the aircraft control action requirements.

If/when an error is projected, a pilot will make the control inputs he/she decides will reduce to a minimum the error remaining after reprojecting the consequences of the corrective action ahead into the next mission increment. The detection of an existing error or of an error emerging from the projected future aircraft status serves to trigger this behavioral process rather than merely informing the pilot of the error or of the amount of error to be eliminated and illustrates the pilot's predominantly looking-ahead orientation (Miller, Galanter and Pribram, 1960). Thus the pilot uses the current state of the aircraft not as the basis for correcting errors if they are present, but instead projects the current state ahead and makes subsequent control inputs based on the projection and what must be done to optimize the future state of the aircraft with respect to remaining mission requirements (Branch, 1978; Dychess, 1977).

THE PILOT'S MISSION PLAN

Having established that a pilot bases aircraft control actions on projections of the aircraft's momentary state, it can be seen that ordinarily he/she is not responding to an external stimulus pattern when a control action response is made. Instead, the pilot responds to the projection of the future state of the aircraft or to an idea, referred to here as a mission cognitive structure or plan, which represents the mission and the aircraft's position in the mission.

Given the assumption of a mission plan to guide control of an aircraft during a flight, it is easily seen that the skilled pilot will apply as many previous experiences as he/she believes may be relevant to the mission being planned. As the plan is developed, a pilot will consider and evaluate an array of potential procedures and maneuvers which can be used to achieve mission goals. From the array of alternatives, the pilot will select those which appear to represent the best mission plan. For example, the pilot may consider duration of the flight, fuel available, weather, load data, takeoff and landing conditions, terrain to be overflown, possible emergency conditions and alternative means for dealing with them. In addition, pilots typically consider other information which appears to be relevant based on their perception of the circumstances of the mission in order to tailor the plan to what they believe is an optimum combination of their capabilities and mission demands.

AIRMANSHIP

The completeness, elaborateness, accuracy and adequacy of the mission plan on which a pilot relies often furnishes an index of his/her experience and flying skill, sometimes referred to as airmanship. The characteristics of a pilot's mission plan also furnish an index of the probable success of the mission. A pilot whose mission plans are relatively complete and comprehensive will be more likely to encounter fewer surprises during a mission, and because of his/her greater preparation, can be expected to deal with such unplanned events or emergencies as may occur with less difficulty than a pilot who is less well prepared.

With the more experienced pilot whose mission plans are relatively complete and comprehensive, it can be assumed that he/she will make appropriate use of the plan to maximize the smoothness, accuracy and eventual success of the flight. Therefore, at any instant during a mission, a skilled pilot may be expected to be dealing with the projection ahead of as many of the alternatives considered in developing the mission plan as he/she believes are applicable (Hyman, 1976).

- Hyman, R. Coding Systems and the Composition of Instructional Materials. AFOSR Report for Contract No. F-44620-73-C-0056. Bolling AFB, DC: Air Force Office of Scientific Research, September 1976.
- Langewieshe, W. Stick and Rudder: An Explanation of the Art of Flying. New York: McGraw-Hill, 1944. Pp25-86.
- Miller, G. A., Galanter, E., and Pribram, K. H. Plans and the Structure of Behavior. New York: Holt, Rinehart and Winston, 1960. Pp 82-93.

THE 1981 NAVAL AND MARINE CORPS AVIATION ANTHROPOMETRY SURVEY AND APPLICATIONS

Harv G. Gregoire, Ph.D.

Naval Air Test Center
Patuxent River, Maryland 20670

The purpose of this survey was to provide a data base of anthropometric information necessary for the design and modification of cockpits, aircrew stations, ejection seats, life support equipment, and aviator flight clothing. The survey was conducted in four phases; (1) a requirements survey and development of a sampling strategy, (2) equipment development, team training, and pilot study, (3) data acquisition (body measurements) at various Naval and Marine Corps Air Stations, and (4) data analysis and reporting. A total of 108 body-size dimensions for males and 112 dimensions for females were measured on 1,087 males and 351 females. Subjects were selected to representatively sample the various subsets of flying populations (i.e., pilot, NFO, enlisted aircrewman, and females). Tabular statistical data is provided for cockpit and aircrew station geometry design applications and also for the design of aircrew-related clothing and body-mounted life support equipment. Additionally, an ancillary anthropometric compatibility program for aviation personnel is briefly discussed.

Introduction

The existing naval aviation anthropometric data used in the design of aircrew station geometry and personal protective and survival equipment was gathered and published in 1964. Since then, naval aviation physical entrance standards have changed and women have recently been authorized to fly as pilots and Naval Flight Officers. These changes in the aviation aircrew population necessitated an updating of previous aviation anthropometric surveys.

The Naval Air Test Center was tasked by the Naval Air Development Center to conduct a large-scale anthropometric survey to provide a data base of over 100 separate body dimensions which would statistically describe the current naval and Marine Corps aviation populations in terms of the extremes, averages, and variations of personnel to be accommodated.

The actual conduct of the survey was preceded by extensive liaison between representatives of government and industry to determine the specific data required by designers of aircraft, flight clothing, and personal equipment. In addition to specific data requirements, the liaison meetings provided valuable insights into the data presentation format judged most useful to designers.

The purpose of the survey was to establish a current statistical body dimension data base which would be valid for use in designing; (1) cockpit and crew station geometry, (2) modifications to current inventory aircraft for aircrew personnel accommodation, (3) ejection seats, (4) flight clothing, and (5) personnel-mounted life support equipment.

Selections, acquisitions, and modifications of the best available anthropometric equipment were made to ensure measurement accuracy and to preclude human error

during the measurement process. Specialized equipment and measuring devices were designed and fabricated in-house.

Discussion

In this survey a sample of 1,087 males and 351 females were preselected by approximate height, weight, and flying duty to accurately represent the anthropometric characteristics and ranges of naval and Marine Corps flying personnel. A total of 156,708 separate measurements of 1,438 total personnel were taken. A comparative analysis between these and previous data was not required or performed.

Since less than 50 females were currently on naval flight status as Pilots or Naval Flight Officers, a sampling strategy was designed which called for selecting a minimum of 350 active duty Navy women within the allowable height and weight parameters specified for aviation personnel.

The survey was conducted to sample a variety of aviation communities (i.e., training, jet, multi-engine, and helicopter, both Navy and Marine). Surveys were conducted at the following locations:

Naval Air Station, Brunswick, Maine

Naval Air Station, Cecil Field, Florida

Naval Air Station, Norfolk, Virginia

Naval Air Station, Oceana, Virginia

Naval Air Test Center, Patuxent River, Maryland

Naval Air Station, Pensacola, Florida

Marine Corps Air Station, Cherry Point, North Carolina

Marine Corps Air Station, El Toro, California

Naval Air Station, San Diego, California

Subsequent to measurement equipment acquisition, a protocol was established to quantify 108 dimensions for males and 112 dimensions for females. Body features were marked with felt-tipped pens to assist anthropometrists in locating physical reference features. Subjects then proceeded through each of the three measuring stations until all measurements were completed. The specific procedure for each measurement is included in reference (a).

Individual identification information was recorded including; sex, age, rank, unit, flight duty, and handedness.

During measurements, male subjects wore gym trunks and female subjects wore non-restrictive, lightweight body suits.

Chronology

<u>Dates</u>	<u>Milestones</u>
25 Oct - 30 Nov 1979	Anthropometric Data Requirements Researched
15 Nov - 15 Jan 1979	*Sampling Plan Designed
20 Nov 1979 thru 20 May 1980	Equipment Acquisition Team Training Pilot Study
1 Oct 1980 thru 15 Apr 1981	Data Acquisition
1 Jun - 1 Sep 1981	Data Analysis - Report Preparation

*A complete description of the sampling strategy is presented in reference (a).

Conclusions

The 1981 Naval and Marine Corps Aviation Anthropometric Survey has been accomplished to provide a current body size data base for engineers and designers tasked with accommodating Naval and Marine Corps aviation personnel. The results are presented in reference (1). Male and female measurements are presented in percentile format; i.e., the respective value on a scale of one hundred that indicates the percent of a distribution that is equal to or below it. These data should be utilized in designing cockpit and crewstation geometry, aircraft personnel accommodation modifications, ejection seats, flight clothing, and personnel-mounted life support equipment.

The effects of flight clothing and equipment as well as postural variables on aircrew accommodation were established in earlier tests (reference (1)). These data have been provided to be used in conjunction with the basic anthropometric data collected for ejection seat cockpit or aircrew station design applications. Additionally, data is provided which quantifies postural variables to be used as dimensional correction data in aircrew design considerations of eye position location and control reach accommodation.

An ancillary anthropometric accommodation coding program has been devised for aviation personnel in the Navy and Marine Corps (references (2) and (3)). The program codes the sitting height, functional reach, buttock-to-knee length, and buttock-leg length with a four digit code which specifies particular ranges of the four respective dimensions. The same four respective dimensions are then coded for each aircraft crewstation in the inventory, i.e., sitting height available, functional reach required to primary flight and emergency controls, buttock-to-knee clearance required for ejection clearance, and leg length necessary to achieve full throw of brakes and rudders. Prior to assignment to particular aircraft, aircrew personnel anthropometric codes are compared with aircraft dimensional codes to prevent assignment to anthropometrically incompatible aircraft.

The 1981 Naval and Marine Corps Aviation Anthropometric Survey data base and ongoing research into specific aircraft anthropometric requirements will be used in continuing efforts to provide the safest and most efficient fit between aircrew, clothing, personnel equipment, and aircraft.

References

1. 1981 Naval and Marine Corps Aviation Anthropometric Survey of 20 Nov 1981
2. OPNAVINST 3710.36A of 10 Jan 1980
3. NAVAIRINST 3710.9 of 19 Sep 1979

DESIGNING FOR COMPATIBILITY

Thomas M. McCloy James A. Boyless Bruce R. Jaeger
Department of Behavioral Sciences and Leadership
United States Air Force Academy

Abstract

Three groups of Air Force Academy cadets--470 freshmen males, 75 freshmen females, and 36 junior human factors engineering students--responded to a questionnaire (Smith, 1981) examining potentially uncertain or confusing display-control relationships. The findings of this study were compared with those from an earlier study (Smith, 1981) which used the same questionnaire. The results indicated strong shared expectations across the two studies for some designs, but a variety of preferences for others. The cadets tended to have more homogenous expectations than the different groups in the Smith (1981) study.

"Probably the designer's most important consideration when designing displays, controls, and workplace layout is to make decisions that are consistent with what the user expects" (Bailey, 1982). Sometimes the display-control relationships for a given user population are so straightforward and natural they form what is commonly called a population stereotype (Smith, 1981). However, in situations where user expectations are not known, the prudent designer should conduct a study to discover these expectations before starting the design process. The present study introduced students in a classroom situation to the concept of designing things for compatibility with the user's expectations.

Method

Subjects

581 Air Force Academy cadets participated in the study, 470 male and 75 female freshmen cadets taking General Psychology, and 36 juniors taking Human Factors Engineering.

Procedure

All participants completed a compatibility questionnaire (Smith, 1981) during one of their normal class periods.

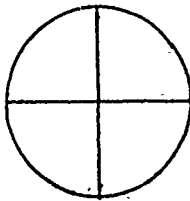
Results and Discussion

In an earlier study, Smith (1981) found some differences in user expectations between three groups--92 male engineering students; 80 females who were friends, relatives, or employees of the engineers; and 55 human factors specialists (mostly males). In the current study, participants were categorized into three groups: freshmen males, freshmen females, and junior human factors students.

Although the primary purpose of the present research was to introduce students to the concept of user expectations in the design of items, certain other issues were also of interest. Would the females respond more like their male peers or like the females in Smith's study? Would the human factors students' preferences differ from student preferences in general? How would this group of users compare to those in Smith's study?

The following five items from one questionnaire are presented with the response frequencies from the present study, as well as those from Smith's study, to allow comparisons. Chi Square statistics were performed on the responses to the present study. Smith also performed Chi Squares and those that were significant are indicated.

1. Quadrant Labels



In what order would you label the 4 quadrants of a circle?

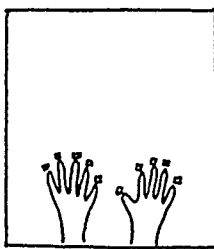
Write in the letters A, B, C, and D, assigning one letter to each quadrant.

Quadrant Label

	Smith <small>p < .001</small>			A F A <small>p < .001</small>		
	E %	F	HFS	M	F	HF
CW from Upper RT	33	26	45	19	25	14
CW from Upper LI	10	11	5	18	7	14
CCW from Upper RT	34	3	5	47	37	31
Reading Order	14	54	43	14	32	42
Other		6	2	1		

In the present study the most popular choices were to letter counter-clockwise from the upper right (congruent with analytic geometry) or to letter in reading order. These were equally popular in Smith's study, although human factors specialists and engineering students also demonstrated a propensity for clockwise from the upper right. Chi Squares were significant at $p < .001$ for both the present ($\chi^2=31.9$) and previous ($\chi^2=67.3$) studies.

2. Numbered Keys



A worker is required to duplicate numbers as they appear on a screen by pressing 10 keys, one for each finger.

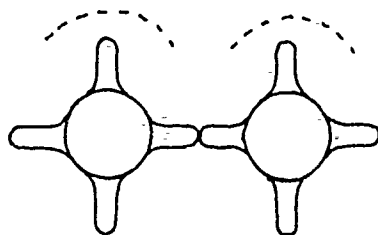
Label the diagram to show how you would assign the 10 numerals to the 10 fingers.

Numbered Keys

	Smith			A F A		
	E %	F	HFS	M	F	HF
Ascending from L to R	70	70	84	59	56	64
Ascending Out from Thumbs	18	16	5	17	29	8
Other	12	14	11	24	15	28

The most popular ordering for both studies was ascending from left to right beginning with zero or one. In the present study the groups were significantly different ($\chi^2=9.7$, $p < .05$).

3. Cross Faucets

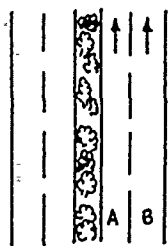


Here are two knobs on a bathroom sink, looking down at them.

Put an arrow on each dotted line to show how you would operate these knobs to turn the water on.

The participants in the present study and those in Smith's indicated significant differences in preference frequencies ($\chi^2=33.3$, $p < .001$ and $\chi^2=21.4$, $p < .01$ respectively). This great diversity in user expectations is probably due to the wide variety of faucet designs found in the commercial marketplace.

4. Highway Lanes



On the 4-lane divided highway pictured here, which is the outside lane?

A _____

B _____

Cross Faucets

		Smith _{people}			AFA _{people}		
		E %	F	HFS	M	F	HF
C	C	17	34	22	22	13	47
C	CC	23	20	13	18	19	14
CC	C	13	26	16	30	54	19
CC	CC	47	20	49	30	14	19

Highway Lanes

		Smith _{people}			AFA _{people}		
		E %	F	HFS	M	F	HF
A		50	51	20	18	14	19
B		50	49	80	82	86	81

All the Air Force Academy students and the human factors specialists from Smith's study strongly preferred the right hand lane, which is the lane farthest from the center of the picture. Smith's groups showed significant preference differences ($\chi^2=13.4$, $p < .01$), with females and engineers showing equal tendencies toward both choices.

5. River Bank



Here is a river, flowing
from East to West.

Is the church on the
— left bank?
— right bank?

River Bank

	Smith			A F A		
	E %	F	HFS	M	F	HF
A	18	16	13	16	18	23
B	82	84	80	84	82	77
Other			7			

The unanimous choice for participants in both studies was to place the church on the right bank, apparently resulting from an orientation of viewing banks looking downstream. However, one student in Smith's study pointed out that hydrologists measure distances along a river from its mouth, thus illustrating a possible conflict between the providers and the users of information.

The results of this study, as well as those from Smith (1981), provide a good introduction to the area of designing for compatibility. Across the five illustrations there were examples of majority agreement representing population stereotypes plus obvious differences in preferred choices. To no great surprise, the three groups of cadets indicated much more homogenous preferences than the groups in Smith's (1981) study.

There is, of course, much room for continued research in this area. Two interesting questions for future study are 1) how do these preferences come to be? and 2) how strong are they, that is, what are the consequences of going against such preferences? For the time being, however, it may be safely said that designers are one step ahead if they consider the user population's preferences while they plan new displays, controls, workplaces, and the like.

References

- Bailey, R.W. Human Performance Engineering. New Jersey: Prentice Hall, 1982.
Smith, S.L. Exploring compatibility with words and pictures. Human Factors, 1981, 23, 305-315.

ATROPINE FAILS TO COUNTER RADIOGENIC BEHAVIORAL DEFICITS

G. Andrew Mickley, Karen E. Stevens, Gerald H. Lawrence
U. S. Air Force Academy, CO 80840

Gregory L. Gibbs and Gerald A. White
Penrose Cancer Hospital, Colorado Springs, CO 80903

ABSTRACT

Exposure of rats to 600 or 1000 rads ⁶⁰Cobalt produces a variety of behavioral symptoms which include aphagia, loss of body weight and reductions in spontaneous locomotor activity. Injections of atropine sulfate (a cholinergic antagonist) fail to reverse these radiogenic changes. Rather, in most cases, atropine treatments heightened the severity of the deficits. Surprisingly, however, a sufficiently large and continuous dose of the drug significantly decreased the lethality of 1000 rads ⁶⁰Co from an LD_{100,30} to an LD_{40,30}. The duality of atropine's long-term and short-term effects brings into question the efficacy of its use to counter military radiation insults.

Introduction

For a number of years the United States Air Force has planned to use anticholinergic drugs (e.g., atropine) as a means to counter some of the behavioral effects produced by chemical warfare agents. With the realization that airmen may find themselves in environments which contain both chemical agents and ionizing radiation, the question has been asked: Will anticholinergic drugs in any way benefit individuals exposed to radiation?

There is, in fact, substantial evidence to suggest that ionizing radiation exposure alters neurotransmitters. Hunt, Dalton, and Darden (1979) for example, noted how exposure to a large dose of high-energy electrons produced a change in both dopamine and acetylcholine systems. The time frame of these neurochemical alterations seemed to correlate well with postirradiation behavioral deficits.

Radiogenic behavioral deficits which have been observed in a variety of species include lethargy, hypokinesia and a decline in motivated and consummatory behaviors (Kimeldorf and Hunt, 1965). Atropine, which blocks cholinergic receptors, reduces cataplexy in dogs (Delashaw, Foutz, Guilleminault and Dement 1979) may have generalized antidepressant effects (Browne, 1979) and attenuates movement disorders (Franz, 1975). The present experiments attempted to discover if atropine might also alleviate the locomotor and consummatory deficits brought about by exposure to ionizing radiation.

Materials and Methods

Subjects

Experimentally naive, male, Sprague Dawley rats (200-400 g) were used

in these experiments Subjects were individually housed in a room illuminated from 6:00 AM to 6:00 PM daily. Ambient temperature was set at 20°C. Purina rat chow and water were continuously available.

Table I
Experimental Groups and Conditions
for Experiment 1

Drug Manipulation ^a	Number of Subjects	Radiation Exposure (⁶⁰ Co)
.4 mg/kg Atropine Sulfate	5	600 R
.4 mg/kg Atropine Sulfate	5	Sham
.8 mg/kg Atropine Sulfate	5	600 R
.8 mg/kg Atropine Sulfate	6	Sham
Saline (.4) ^b	5	600 R
Saline (.4) ^b	5	Sham
Saline (.8) ^c	5	600 R
Saline (.8) ^c	5	Sham
.8 mg/kg Atropine Sulfate	6	1000 R
.8 mg/kg Atropine Sulfate	6	Sham
1.056 mg Atropine Sulfate	5	1000 R
1.056 mg Atropine Sulfate	6	Sham
Saline (.8) ^c	6	1000 R
Saline (.8) ^c	6	Sham
Sham Osmotic Pump Implant	5	1000 R
Sham Osmotic Pump Implant	5	Sham

a. Indicated is half the total daily dose. All groups, except 1.056 mg atropine sulfate, received the noted dose twice a day (7:30 AM and 4:00 PM). The 1.056 mg atropine sulfate is a total daily dose which was administered continuously, 24 hours a day, at the rate of 44 µg/hour.

b. Volume equal to that administered to .4 mg/kg atropine sulfate group.

c. Volume equal to that administered to .8 mg/kg atropine sulfate group.

Experimental Groups and Procedures

Forty-two rats were exposed to either 600 rads or 1000 rads ⁶⁰Co (see irradiation procedure). Forty-four others were sham irradiated (see Table I). Most subjects received injections (.4 mg/kg or .8 mg/kg) of atropine sulfate or saline, i.p., five minutes before irradiation. In these animals equivalent follow-up injections were administered twice a day (7:30 AM and 4:00 PM) during the next two days (5 injections total).

The morning of the day of irradiation exposure other rats were chronically implanted with an osmotic minipump (Alzet) which delivered atropine sulfate s.c. continuously at a rate of 44 µg/hour (1.056 mg/day) for one week. Surgery was accomplished under methoxyflurane and the animals were allowed to recover from this anesthesia before measurements were taken.

Measurement of Dependent Variables

Locomotor Activity

Spontaneous locomotor activity was measured using one of two apparatuses. All animals exposed to 600 rads ⁶⁰Co (and sham controls) were individually placed in a Lafayette Instrument's Animal Activity Monitoring System two days before irradiation. Immediately after this baseline period the rats were irradiated and then replaced in the activity monitoring cages. In this postirradiation test period locomotor movements were recorded continuously for five days.

The animals exposed to 1000 rads ⁶⁰Co (and their sham controls) followed the same procedure as described above but locomotor activity was recorded via Columbus Instrument's, Automex II system.

Consummatory Behaviors and Body Weight

Food consumption was measured daily for two days prior to irradiation (baseline) and then for 30 days immediately postirradiation. Daily body weights were also noted in this time frame.

⁶⁰Co Radiation Exposure

Rats were placed into individual constraining polyethylene tubes immediately before they were irradiated. The tops of the tubes were covered by a thin (.5mm) lead sheet.

Subjects received a whole-body midline tissue dose of either 600 or 1000 rads ⁶⁰Co. Time of exposure was 4.41 to 4.52 minutes for 600 rads, and 7.78 to 8.05 minutes for 1000 rads. Therefore, within the time of collection of the results reported, the midbody dose rate decayed from 136.05 to 132.74 rads/min for 600 rads, and, during a different time period, 128.53 to 124.22 rads/min for 1000 rads. The source to subject distance was 80 cm and the field of exposure varied from 15x15 cm (600 rads) to 25x25 cm (1000 rads). Radiation originated above the animal from a Theratron 80 irradiator manufactured by Atomic Energy of Canada.

The ^{60}Co source was a single sphere of approximately 4580 curies (as of December 1980). It was driven in a horizontal direction by air pressure in the Theratron 80.

Dosimetry was accomplished in air with ionization chambers (Spokas Model 1000) whose calibration is traceable to the National Bureau of Standards. The chambers were equipped with a ^{60}Co build-up cap. The doses and back-scatter factors were computed from Depth/Dose Tables (Cohen, Jones and Greene, 1972).

Sham irradiated animals were transported to the radiation facility and placed in an irradiation tube for a time equal to that of the actual radiation exposure.

Results

All postirradiation measures (spontaneous locomotor activity, food consumption and body weight) were converted to Difference Scores (DS) reflecting changes from an average preirradiation baseline. These data were analyzed using a $2 \times 2 \times 2$ analysis of variance for each of the two doses of radiation with repeated measures over time:

- (1) 600 rads; [irradiation (gamma/sham x drug (atropine/saline) x drug dose (.4 mg/kg / .8 mg/kg)] .
- (2) 1000 rads; [irradiation (gamma/sham x drug (atropine/saline) x drug dose (.8 mg/kg / 1.056 mg)] .

Locomotor Activity

Two blocks of time were analyzed for both doses of radiation: average changes from baseline activity on days 1-2 and days 3-5 postirradiation.

Both 600 and 1000 rads ^{60}Co irradiation produce a severe and statistically significant decline in spontaneous locomotor activity (600 rads: $p < .05$, $F = 4.828$, $df = 1,33$; 1000 rads $p < .01$, $F = 99.053$, $df = 1,37$). In the rats exposed to 600 rads neither dose of (.4 mg/kg or .8 mg/kg) of atropine sulfate significantly altered this reduction ($p > .05$, $F = 0.310$, $df = 1,33$). However, in subjects exposed to 1000 rads a significant drug effect is present ($p < .05$, $F = 4.348$, $df = 1,37$). Contrary to the working hypothesis, rats treated with 1.05 mg atropine daily were significantly hypokinetic (Tukey's HSD test, $p < .05$) when compared to irradiated rats treated with saline (see figure 1). The analysis of variance revealed a significant drug/dose interaction in the 1000-rad animals. This suggested that while increased control injection (saline) doses had no effect on locomotor activity, increased doses of atropine tended to lower the quantity of spontaneous movements ($p < .01$, $F = 8.869$, $df = 1,37$). It should be noted that sham irradiation did not significantly decrease locomotion in this experiment (sign tests $p > .05$) and that atropine treatments didn't alter the locomotor activity of these control animals (sign tests $p > .05$).

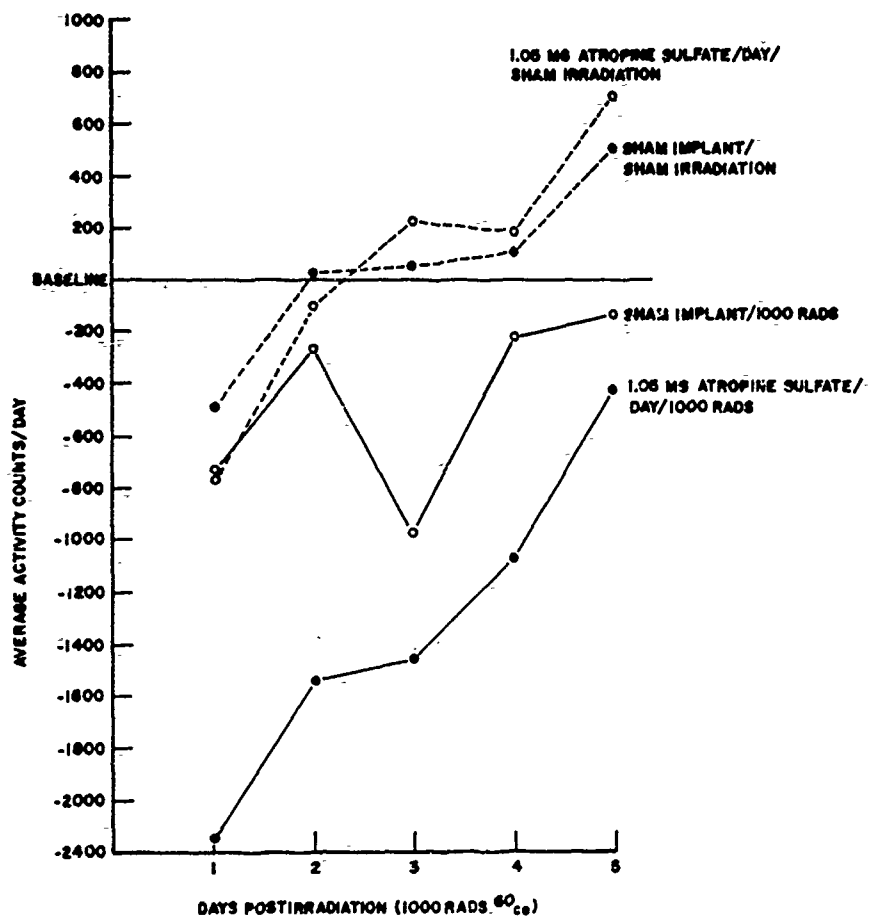


Figure 1. Average change in spontaneous locomotor activity after 1000 rads ^{60}Co or sham irradiation. Some rats were implanted with an osmotic pump which supplied a continuous (44 $\mu\text{g}/\text{hour}$) dose of atropine sulfate equal to 1.05 mg/day. Other animals underwent sham implant procedures.

Food Consumption

These data were analyzed using an analysis of variance technique as previously described. Data were combined and averaged over four periods postirradiation (1-5, 6-10, 11-20, and 21-30 days). There was a significant decline in food consumption in irradiated subjects (600 rads: $p < .01$, $F = 22.684$, $df = 1,33$; 1000 rads: $p < .01$, $F = 61.895$, $df = 1,37$). This deficit was most prominent during the initial days postirradiation. While there was no statistically significant drug effect present in these data (600 rads: $p > .05$, $F = 0.110$, $df = 1,33$; 1000 rads $p > .05$, $F = 1.144$, $df = 1,37$) atropine treatments of all doses tended to lower food consumption (when compared to saline injected irradiated rats) especially on the days immediately following irradiation. Food consumption was not lessened by sham irradiation or by atropine treatment alone ($p > .05$ sign tests).

Body Weight

Body weight changes from baseline were averaged over four blocks of time: 1-5, 6-10, 11-20, and 21-30 days postirradiation. These data were analyzed using an analysis of variance with repeated measures over time. A statistically significant drop in body weight occurred in irradiated subjects (600 rads: $p < .01$, $F = 59.919$, $df = 1,33$; 1000 rads $p < .01$, $F = 31.482$, $df = 1,37$) in the first days after exposure. Once again, atropine does not significantly alter body weight when compared to saline injected or sham-implanted controls ($p > .05$). Still, especially in animals which received continuous infusion of atropine (1.05 mg/day) the trends suggest that the drug lessened weight. This phenomena does not appear to be specific to irradiated animals. Atropine also reduces (albiet not significantly) weight gains of sham irradiated control subjects.

Lethality

All animals irradiated with 600 rads ⁶⁰Co survived 30 days postirradiation. However, this was not the case in animals exposed to 1000 rads. All of the saline-injected, irradiated rats died within 30 days. This was also the case for subjects injected with 8 mg/kg atropine twice daily.

Although atropine seems to cause deficits in almost every behavioral measure observed in these studies, a sufficiently large dose of the drug given continuously apparently reduces lethality (see figure 2). Those rats which received 1.05 mg daily doses of atropine continuously for a week after exposure exhibited an $LD_{40,30}$ (40% of the animals dying within 30 days) as compared to the $LD_{100,30}$ after sham drug treatments.

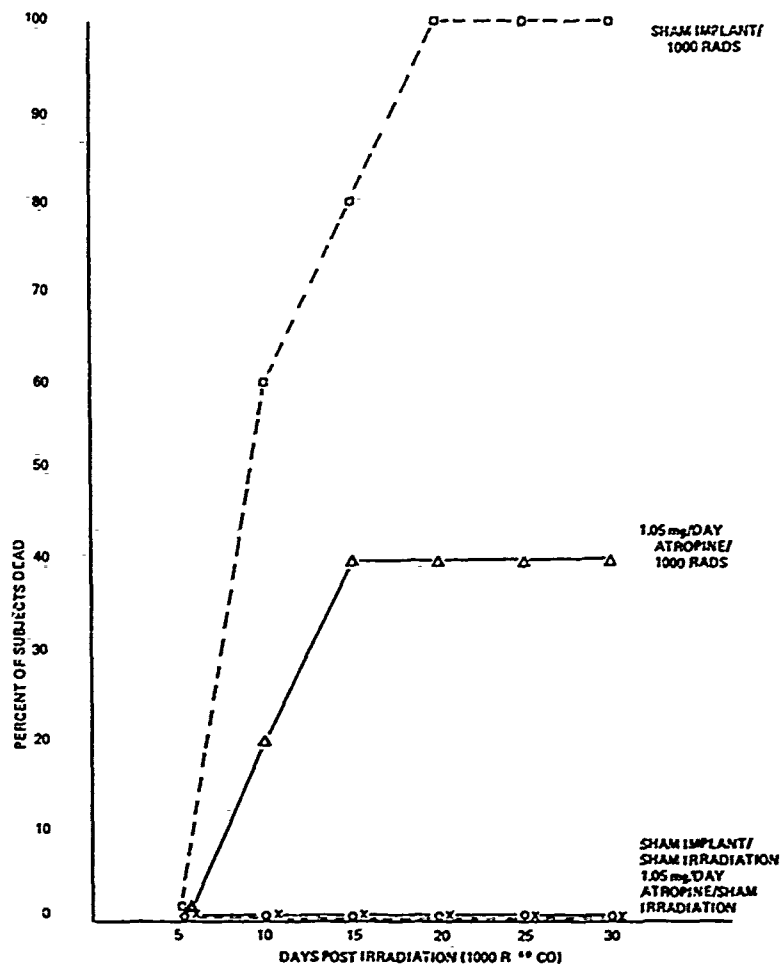


Figure 2. Lethality in subjects exposed to 1000 rads ⁶⁰Co or sham irradiation. All sham exposed subjects lived the entire 30 days. Irradiated rats given 1.05 mg/day continuously (44 µg/hour) did not die in numbers as great (LD 40,30) as those which underwent sham implant procedures (LD 100,30).

This difference is statistically significant ($p = .031$, sign test). Atropine may have some long-term radioprotective effects which contrast dramatically with its short-term behavioral decrements.

Discussion

When rats were exposed to either 600 or 1000 rads ⁶⁰Co they experienced a decrement in a variety of behaviors. Although the animals exposed to the larger dose experience greater deficits, both irradiations produced reliably lower spontaneous locomotor activity, body weight, and eating behaviors. Contrary to the working hypothesis of this study, atropine treatment did not benefit these animals. Often the drug actually enhanced (albeit not significantly) the severity of the deficits described. These findings are in consonance with a study by Smith and Morris (1963) which suggests that atropine does not reduce taste aversions produced by exposure to ionizing radiation.

The data reflected in this report are not consistent with the literature reviewed in the introduction which suggested that muscarinic receptor blockage might alleviate movement disorders (Delashaw, Foutz, Guillemineault, and Dement 1979; Browne, 1979; Franz, 1975). Still, the data do agree with an older body of work which notes that atropine can cause drowsiness and depress the EEG arousal response (Ostfeld and Arguete, 1962). In addition, atropine or scopolamine can disrupt behavioral responses at the same time that the EEG is altered. Conversely, physostigmine (a cholinomimetic) restores both electrophysiological and behavioral responses to normal (Toman and Davis, 1949; Longo, 1966) and has also been known to benefit certain movement disorders (Davis, Hollister, Barchas and Berger, 1976; Klawans and Rubovits, 1974). This literature would suggest, however, that atropine injections might produce behavioral deficits in sham irradiated animals. But this was not observed in the present studies. Atropine did seem to enhance one very important measure: Survival. There appears to be a significant increase in the number of

atropine-treated rats which survived 1000 rads ⁶⁰Co radiation for 30 days. This was only true if a large (1.05 mg/day) dose of atropine was administered continually over a long period of time.

The present study employed a species which is in many ways dissimilar from man. These results should certainly be considered preliminary to any decision concerning the efficacy of the use of atropine in wartime radiation environments. However, the contrast between the apparent short-term (behaviorally unfavorable) and long-term (survivability) effects of the drug raises a variety of questions to be answered by future experiments on higher mammals and primates.

REFERENCES

- Browne, R. G. "Effects of antidepressants and anticholinergics in a mouse behavioral despair test." European Journal of Pharmacology, 1979, 58, 331-334.
- Cohen, M., Jones, D. E. A. and Greene, D. (Eds), "Central Axis Depth Dose Data for Use in Radiotherapy." British Journal of Radiology, Suppl. 11, 1972.
- Davis, K. L., Hollister, L. E., Barchas, J. D., and Berger, P. A. "Choline in tardive dyskinesia and Huntington's disease." Life Science, 1976, 19, 1507-1516.
- Delashaw, J. B., Foutz, A. S., Guilleminault, C., and Dement, W. C. "Cholinergic mechanisms and cataplexy in dogs." Experimental Neurology, 1979, 66, 745-757.
- Franz, D. N. "Drugs for parkinson's disease; centrally acting muscle relaxants" in The Pharmacological Basis of Therapeutics (L. S. Goodman and A. Gilman, Eds.) New York, Macmillan Publishing Co., Inc., 1975.
- Hunt, W. A., Dalton, T. K., and Darden, J. H. "Transient alterations in neurotransmitter activity in the caudate nucleus of rat brain after a high dose of ionizing radiation." Radiation Research, 1979, 80, 556-562.
- Kimeldorf, D. J. and Hunt, E. L., Ionizing Radiation: Neural Function and Behavior. New York/London, Academic Press, 1965.
- Klawans, H. L. and Rubovits, R. "Effects of cholinergic and anticholinergic agents on tardive dyskinesia." Journal of Neurology, Neurosurgery and Psychiatry, 1974, 27, 941-947.
- Longo, V. G. "Behavioral and electroencephalographic effects of atropine and related compounds." Pharmacology Review, 1966, 18, 965-996.
- Ostfeld, A. M. and Arguete, A. "Central nervous system effects of hyoscine in man." Journal of Pharmacology and Experimental Therapeutics, 1962, 137, 133-139.
- Smith, J. C. and Morris, D. D. "Effects of Atropine Sulfate on the Conditioned Aversion to Saccharin Fluid with X-Rays as the Unconditioned Stimulus." Radiation Research, 1963, 28, 186-190.
- Toman, J. E. P. and Davis, J. P. "The effects of drugs upon the electrical activity of the brain." Pharmacology Review, 1949, 1, 425-492.

This research was funded by the School of Aerospace Medicine, Brooks Air Force Base. Research Contract Number SAM TSO 80-3.

REFERENCES

- Browne, R. G. "Effects of antidepressants and anticholinergics in a mouse behavioral despair test." European Journal of Pharmacology, 1979, 58, 331-334.
- Cohen, M., Jones, D. E. A. and Greene, D. (Eds), "Central Axis Depth Dose Data for Use in Radiotherapy." British Journal of Radiology, Suppl. 11, 1972.
- Davis, K. L., Hollister, L. E., Barchas, J. D., and Berger, P. A. "Choline in tardive dyskinesia and Huntington's disease." Life Science, 1976, 19, 1507-1516.
- Delashaw, J. B., Foutz, A. S., Guilleminault, C., and Dement, W. C. "Cholinergic mechanisms and cataplexy in dogs." Experimental Neurology, 1979, 66, 745-757.
- Franz, D. N. "Drugs for parkinson's disease; centrally acting muscle relaxants" in The Pharmacological Basis of Therapeutics (L. S. Goodman and A. Gilman, Eds.) New York, Macmillan Publishing Co., Inc., 1975.
- Hunt, W. A., Dalton, T. K., and Darden, J. H. "Transient alterations in neurotransmitter activity in the caudate nucleus of rat brain after a high dose of ionizing radiation." Radiation Research, 1979, 80, 556-562.
- Kimeldorf, D. J. and Hunt, E. L., Ionizing Radiation: Neural Function and Behavior. New York/London, Academic Press, 1965.
- Klawans, H. L. and Rubovits, R. "Effects of cholinergic and anticholinergic agents on tardive dyskinesia." Journal of Neurology, Neurosurgery and Psychiatry, 1974, 27, 941-947.
- Longo, V. G. "Behavioral and electroencephalographic effects of atropine and related compounds." Pharmacology Review, 1966, 18, 965-996.
- Ostfeld, A. M. and Arguete, A. "Central nervous system effects of hyoscine in man." Journal of Pharmacology and Experimental Therapeutics, 1962, 137, 133-139.
- Smith, J. C. and Morris, D. D. "Effects of Atropine Sulfate on the Conditioned Aversion to Saccharin Fluid with X-Rays as the Unconditioned Stimulus." Radiation Research, 1963, 28, 186-190.
- Toman, J. E. P. and Davis, J. P. "The effects of drugs upon the electrical activity of the brain." Pharmacology Review, 1949, 1, 425-492.

This research was funded by the School of Aerospace Medicine, Brooks Air Force Base. Research Contract Number SAM TSO 80-3.

exposure were examined and a cholinergic antagonist, Atropine Sulfate, was tested for reversal of any of the observed changes. Atropine is currently used to counter the effects of certain chemical warfare agents and due to its effect on the cholinergic system, it could provide some measure of protection against radiation produced imbalances in that neurotransmitter system.

Materials and Methods

Subjects

Twenty-four experimentally naive male Sprague-Dawley rats (200-400 g) were used in these experiments. Subjects were individually housed in a room illuminated from 6:00 AM to 6:00 PM daily. Ambient temperature was set at 20°C. Purina rat chow and water were continuously available.

Experimental Groups and Procedures

Twelve of the subjects were exposed to 600 rads Neutrons (see irradiation procedures) and the other twelve were sham irradiated. Half of the subjects in each one of these groups received .8 mg/kg atropine sulfate i.p. five minutes before radiation exposure or sham irradiation. The other half received an equal volume of saline, i.p. Follow-up injections of the same dosage were administered twice a day (7:30 AM and 4:00 PM) during the next two days (5 injections total).

Measure of Dependant Variables

Spontaneous locomotor activity was measured in a Lafayette Instrument's Animal Activity Monitoring System. Activity was recorded for two days prior to and five days after irradiation or sham.

Food and water consumption were measured for two days prior to irradiation (baseline) and then for 30 days postirradiation. Daily body weights were also noted in this time frame.

In addition, systolic blood pressure was taken on the rats. An IITC indirect blood pressure apparatus was used, which consists of a 3.4 cm inflatable tail cuff with built in photoelectric sensor to record pulses. The cuff was inflated with an electric automatic pump in order to insure consistency. This technique is advantageous in that it does not require heating the rat as do other indirect methods.

Neutron Irradiation Procedure

Rats received 600-rad whole-body neutron exposures. A typical dose rate was 21 rads/minute. These irradiations were accomplished using a Kaman Sciences Corporation Model A-1254 neutron generator which produces 14 Mev neutrons from the Deuterium-Tritium reaction. A typical exposure run used 170kv potential, a beam current of 1.3 ma and a target current of 0.55 ma. This produced a flux of 5×10^7 n/cm² sec on centerline at 5 cm.

The flux was measured using a modified form of the "Texas Convention" (DeSeote, Gijbels and Hoste, 1972). Copper foils (1.3 cm diameter,

0.159 grams mass) were activated using the $^{63}\text{Cu} (n, 2n) ^{62}\text{Cu}$ reaction which has a 10.8 mev threshold (Alfev, Drynkin, Leipunskaya and Kasatkin, 1970). The cross section varies but a value of 500 millibarns was used (as in the "Texas Convention"). The ^{62}Cu decays by β^+ emission and its activity was calculated from a determination of the count rate from the 0.511 Mev annihilation photons. The copper foils were encased in 1 cm of plastic on all sides to capture the β^+ emissions. The plastic-copper sandwich was counted on a 7.8 x 7.8 cm NaI detector. Due to the large fluences, the sandwich was counted at 10.5 cm from the detector.

Activity of the ^{62}Cu was calculated by reference to a 2% standard ^{24}Na source in the same geometry. The counts were taken for one minute, starting two minutes after the irradiation ended. A typical irradiation lasted 30 minutes and produced a count rate of 20,000 counts/minute.

The midline absorbed dose to the rat was determined from the fluence by using previously published calculations (NCRP Report Number 38, 1971). A value of 7×10^{-9} rads/cm² was used assuming 2 cm depth and neutron effects only. Total gamma plus neutron would have given a value of 7.5×10^{-9} .

By using a combination of timing and an external BF_3 detector, each rat was given a dose of 600 ± 10 rads. Before each run a foil was placed on the center side of the rat cage to determine the dose received. Rats were irradiated 5 cm from the target. At this distance the flux was reduced to 85% of its centerline value at points 2 cm to either side of center.

Results

Locomotor Activity

Locomotor activity was recorded 24 hours a day and average activity counts per hour were calculated for each day activity was monitored. For statistical analysis, difference scores were computed from the activity/hour data: Average daily postirradiation locomotor activity/hour minus average daily preirradiation baseline locomotor activity/hour = DS. These data were analyzed using a 2 x 2 analysis of variance [irradiation (neutron/sham) x drug (.8 mg/kg atropine/saline)] with repeated measures over time. Two blocks of time were analyzed in this manner: days 1-2 and days 3-5 postirradiation.

Although there was a trend toward hypokinesia, the analysis revealed that 600 rads neutron exposure did not significantly reduce locomotor activity when compared to sham irradiated controls ($p > .05$, $F = 3.234$, $df = 1,20$). Likewise, atropine injections did not significantly alter locomotor activity ($p > .05$, $F = .675$, $df = 1,20$). In addition, a sign test revealed that sham irradiated rats did not significantly ($p > .05$) reduce their locomotor activity from baseline levels.

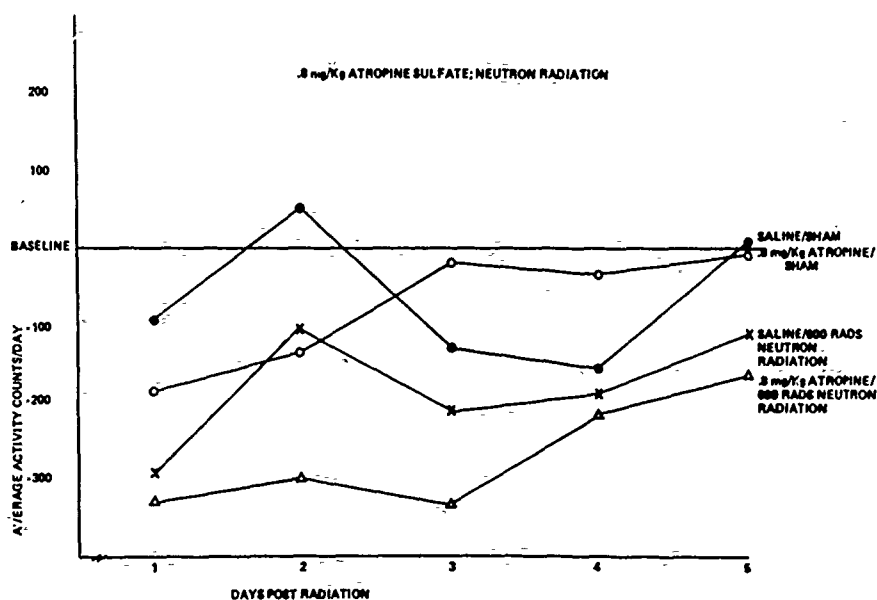


Figure 1

Average change in spontaneous locomotor activity after 600 rads neutrons or sham irradiation. Some rats received .8 mg/kg atropine sulfate; the others were injected with an equal volume of saline.

Food Consumption

Data was recorded and converted to difference scores (change in food consumed as compared to baseline). These numbers were then averaged over four separate periods postirradiation: 1-5, 6-10, 11-20, 21-30 days. The average data comprised the repeated measure variable in the analysis of variance (previously described above).

Although there was an initial deficit in food consumption, neutron radiation did not produce a significant difference in this measure ($p > .05$, $F = 0.254$, $df = 1,20$). Animals seemed to recover from the aphagia in about four days regardless of whether they received atropine or saline injections. Atropine injections actually seemed to reduce food consumption although this was not a statistically significant decrement ($p > .05$, $F = 0.171$, $df = 1,20$). During the first period (days 1-5) after sham irradiation, atropine or saline-treated rats did not consume quantities of food which were significantly different from baseline ($p > .05$, sign tests).

Body Weight

Data were collected and difference scores computed as described previously.

Body weight changes from baseline were averaged over four blocks of time: 1-5, 6-10, 10-20, and 21-30 days after neutron irradiation. These data were used to compute the analysis of variance as described earlier. The neutron data indicate that there was a significant change in body weight after irradiation ($p < .01$, $F = 14.244$, $df = 1,20$). There was also a most significant effect across time ($p < .01$, $F = 338.477$, $df = 3,60$). This overall change reflected an initial decrement in weight in the irradiated groups during days 1-5 ($p < .05$, Tukey's HSD). However, the decline was followed by a recovery in which all groups of rats are eventually (days 21-30) significantly heavier than baseline ($p < .05$, sign tests).

Despite the fact that neutron irradiated rats which received atropine showed greater initial deficits than those which received saline, there was no significant overall drug effect ($p > .05$, $F = 0.005$, $df = 1,20$).

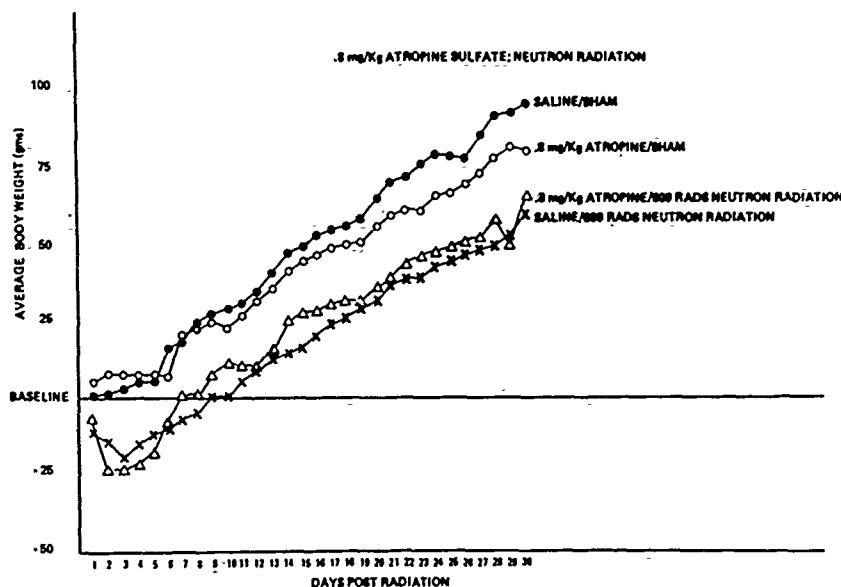


Figure 2

Average change in body weight after 600 rads neutrons or sham irradiation. Some rats received .8 mg/kg atropine sulfate; the others were injected with an equal volume of saline.

Water Consumption

Data was recorded and converted to difference scores as described earlier (change in water consumed as compared to baseline). For the

present analysis we used water consumption difference scores at five specific days postirradiation: 1, 5, 10, 20, and 30. These data were analyzed using the analysis of variance as described previously.

Sham irradiation failed to significantly alter drinking scores ($p > .05$, sign tests, days 1,5). Although there was an initial deficit in drinking ($p < .05$, sign test; day 1) after neutron exposure this hypodipsia was also not a significant one over the longer term ($p > .05$, $F = 2.481$, $df = 1,20$).

Atropine did not significantly reduce drinking ($p > .05$, $F = 1.245$, $df = 1,20$); but there was a trend in this direction.

Blood Pressure

Systolic blood pressures were taken immediately before irradiation and then one and two days after exposure to 600 rads neutrons. Numbers used for statistical analysis consisted of the mean of at least three blood pressure readings. Difference scores (DS) were computed from these data: $DS = \text{postirradiation blood pressure} - \text{preirradiation baseline}$. These data were analyzed using a 2×2 analysis of variance [Irradiation (neutrons/sham) \times drug (.8 mg/kg atropine/saline)].

Overall, animals which were irradiated were slightly hypotensive during the two days after exposure to neutrons (average decrease of 6.625 mm Hg) while sham irradiated subjects were hypertensive (average increase of 12.092 mm Hg) in this same period. There was a significant difference in blood pressure due to irradiation ($p < .05$, $F = 4.884$, $df = 1,20$).

Atropine injections caused no significant change in systolic pressure ($p > .05$, $F = .073$, $df = 1,20$).

Unfortunately the average preirradiation baseline blood pressures were quite variable between groups. This fact may limit the generalizability of the data.

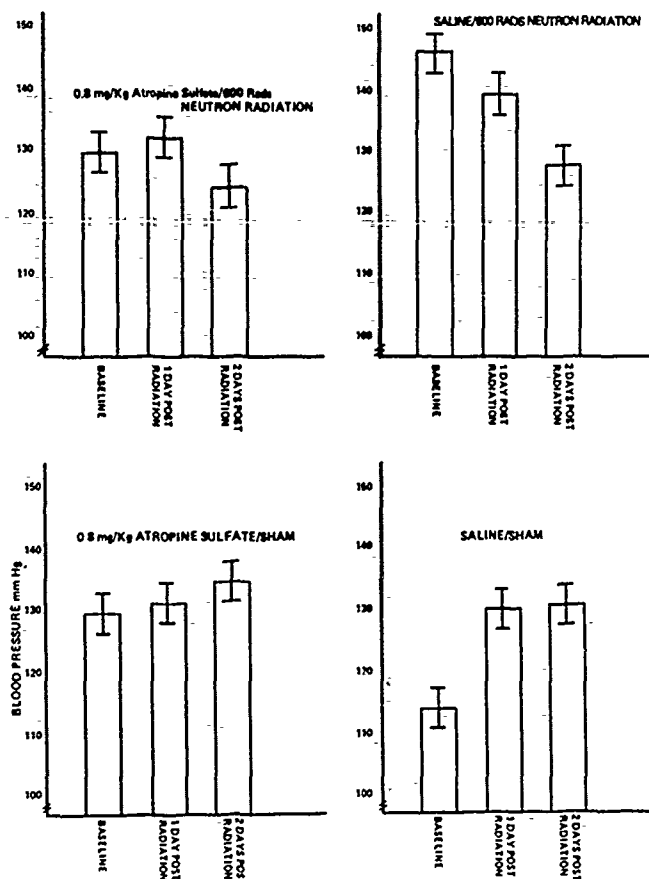


Figure 3

Baseline and postirradiation (days 1, 2) blood pressures of rats exposed to 600 rads neutrons of sham irradiation. Half the animals received .8 mg/kg atropine sulfate; the others were injected with an equal volume of saline. Variance is indicated by the standard error of the mean.

Lethality

All animals irradiated with 600 rads neutron survived 30 days postirradiation. Rats were sacrificed after this time.

Discussion

Neutron exposure initially reduced all the measures reported here; however, not all deficits were statistically significant. These data are in consonance with those reported by Young and Kunihiro (Young and Kunihiro, 1980) which suggest that neutron fields do produce behavioral changes but the changes are not as great as those produced by other ionizing radiation fields; likewise, others (George, Chaput, Verrelli and Barron, 1971) have found that pigs exposed to other forms of ionizing radiation are more likely to suffer an early transient incapacitation

than are those subjects irradiated with neutrons. Six hundred rads gamma irradiation produces greater behavioral deficits than those caused by an equivalent dose of neutrons (Mickley, Stevens, Lawrence, Gibbs, and White, in press).

Although little is known about the physiological mechanisms behind these differences in postirradiation performance, Catravas and McHale (Catravas and McHale, 1973) have reported radiation-specific neurotransmitter enzyme changes. When rats received radiation rich in neutrons a pronounced decrease in the activity of monoamine oxidase occurred (probably increasing catecholamine levels); whereas, when the animals were exposed to the same dose of radiation rich in gamma rays, the activity of this enzyme was markedly increased (probably decreasing catecholamines). Perhaps the neurotransmitter changes produced by gamma exposure make a more profound impact on behavior than do those produced by neutrons.

Cholinergic blockade by atropine failed to reduce the mild deficits in behavior which were observed after neutron exposure. There was a slight trend which suggests that atropine may actually enhance the effects of neutron irradiation. It may be the case, then, that neutrons produce physiological events more closely akin to cholinergic antagonism than to cholinomimetic actions.

REFERENCES

- Kimeldorf, D. J. and Hunt, E. L. Ionizing Radiation; Neural Function and Behavior, Academic Press, New York/London, 1965.
- Mickley, G. A. and Teitelbaum, H. "Persistence of lateral hypothalamic-mediated behaviors after a supralethal dose of ionizing radiation." Aviat. Space Environ. Med. 49, 868-873 (1978).
- Stricker, E. M. and Zigmond, M. J. "Brain Catecholamines and the lateral hypothalamic syndrome" in Hunger: Basic Mechanisms and Clinical Applications (D. Novin, W. Wyrwicka and G. Bray, Eds.), pp. 19-32 Raven Press, New York, 1976.
- Marshall, J. F., Richardson, J. S., and Teitelbaum, P. "Nigrostriatal bundle damage and the lateral hypothalamic syndrome." J. Comp. Physiol. Psychol., 87, 808-830, (1974).
- Mickley, G. A. and Teitelbaum, H. "Movement induced in cataleptic rats: Differential effects produced by stimulation of the lateral hypothalamus, substantia nigra and reticular formation," Psychopharmacol. 57, 145-149, (1978).
- Hunt, W. A., Dalton, T. K., and Darden, J. H. "Transient alterations in neurotransmitter activity in the caudate nucleus of rat brain after a high dose of ionizing radiation," Radiat. Res. 80, 556-562, (1979).
- Varagic, V. Stepanovic, S. Svecenski, N., and Hajdukovic, S. "The effect of X-irradiation on the amount of catecholamines in heart atria and hypothalamus of the rabbit, and in brain and heart of the rat," Int. J. Rad. Biol., 12, 113-119, (1967).

- Johnsson, J. E., Owman, C. H., and Sjöberg, N. O., "Tissue content of noradrenaline and 5-hydroxytryptamine in the rat after ionizing radiation", Int. J. Radiat. Biol., 18, 311-316, (1970).
- Palaic, D. J. and Supek, Z., "Liberation of brain 5-hydroxytryptamine and noradrenaline by X-ray treatment in the newborn and adult rat", J. Neurochem., 13, 705-709, (1966).
- Hunt, W. A., and Dalton, T. K., "Reduction in cyclic nucleotide levels in the brain, after a high dose of ionizing radiation", Radiat. Res., 83, 210-215, (1980).
- Catravas, G. N. and McHale, C. G., "Effects of mixed gamma-neutron radiation on the activity of brain enzymes involved in neurotransmitter metabolism", AFRRI Scientific Report SR 73-17, Armed Forces Radiobiology Research Institute, Bethesda, MD, 1973.
- DeSeote, D., Gijbels, R., and Hoste, J. "Neutron Activation Analysis," Vol. 34 in Chemical Analysis; A Series of Monographs on Analytical Chemistry and its Applications (P. J. Elving and I. M. Kolthoff, Eds.), pp. 76-80, Wiley-Interscience, New York, 1972.
- Alrev, A. I., Drynkin, V. I., Leipunskaya, D. I., and Kasatkin, V. A., Handbook of Nuclear Data for Neutronactivation Analysis, (B. Benny, Trans.), pp. 143-144, Ketter Press, Jerusalem, 1970.
- NCRP Report Number 38, Protection Against Neutron Radiation, National Council on Radiation Protection and Measurements, Washington, D.C., 1971.
- Young, R. W. and Kunihiro, D. M., "The relative effectiveness of high energy and fission spectrum neutrons in producing behavioral incapacitation, emesis and mortality in the monkey (Abstract EA-4)". Paper presented at the 28th Annual Scientific Meeting of the Radiation Research Society (1980) and published in the abstracts, p. 53.
- George, R. E., Chaput, R. L., Verrelli, D. M., and Barron, E. L., "The relative effectiveness of fission neutrons for miniature pig performance decrement", AFRRI Scientific Report SR71-2, Armed Forces Radiobiology Research Institute, Bethesda, MD, 1971.
- Mickley, G. A., Stevens, K. E., Lawrence, G. H., Gibbs, G. L. and White, G. A., "Atropine fails to counter radiogenic behavioral deficits", paper presented at the 8th Psychology in the Department of Defense Symposium, (In press).
- Catravas, G. N. and McHale, C. G., "Activity changes of brain enzymes in rats exposed to different qualities of ionizing radiation", AFRRI Scientific Report SR 73-19, Armed Forces Radiobiology Research Institute, Bethesda, MD, 1973.

This research was funded by the School of Aerospace Medicine, Brooks Air Force Base. Research Contract Number SAM TSO 80-3.

CAUSAL ANALYSIS OF MILITARY COLD INJURIES

James B. Sampson

and

Jared B. Jobe

US Army Research Institute of Environmental Medicine
Natick, Massachusetts 01746

CAUSAL ANALYSIS OF MILITARY COLD INJURIES

ABSTRACT

The present study investigated important antecedent events leading to cold injuries during two recent cold weather maneuvers. Structured interviews of cold injured patients were conducted and an analysis of precipitating factors was performed to identify major events leading to injury. The results show a common sequence of events including physical over-exertion followed by prolonged immobility. Equipment failures and personal medical histories were also implicated as factors. Discussion is given to overall contribution of human decision-making at the personal, group and command/leadership levels.

CAUSAL ANALYSIS OF MILITARY COLD INJURIES

When a large number of military personnel are required to operate in extreme cold weather, cold injuries inevitably occur in spite of the best preparation. These injuries are the result of a complex array of events, decisions and environmental conditions which are not under complete human control. Nevertheless, proper preparation by military units can do a great deal to reduce the potential hazards and number of serious injuries. Although the injury is the ultimate result of exposure of tissue to cold, it is the conditions leading to exposure that need investigation. The cause of exposure involves actions and equipment which are under some degree of control by the men and women involved.

Research is required to better understand the more common routes to injury. Such research involves careful and accurate observation and recording of events, decisions and equipment failure which occur just prior to injury. The present report is based on a series of survey studies designed to assess a broad spectrum of medical problems which occur during field exercises in extreme weather. The presentation will focus on cold injuries and an explanation of their causes. This analysis is based on a survey involving careful recording and classification of the type of injuries as they were typically reported to the medical system during two recent cold weather military exercises. In addition to medical complaints, the survey documented operational events, equipment employed and daily climatic conditions.

METHOD

Medical data were collected at all available medical treatment facilities of two separate cold weather exercises, involving approximately 10,000 personnel in each operation. The first exercise occurred in upper New York in January 1980; whereas, the second occurred in Alaska in January 1981.

As part of the overall survey, structured interviews were obtained from each individual identified by the Preventive Medicine Officer as a person with a possible cold injury. These patients were asked, by trained interviewers, to describe their injury, accompanying signs and symptoms, the circumstances surrounding the injury, the operational events, the state of their equipment and clothing, their level of training and prior experience, their health status, their smoking and drinking habits as well as information on gender, rank and job duties.

Case histories were written for each interview describing the individual, the injury, the circumstances and an estimate of probable contributing factors. A summary list of all unique terms used to describe precipitating factors was made by members of the research team. The terms were then grouped under the general headings of Operational Factors, Equipment Factors, Human Factors and Accident Factors. The specific elements of each factor are as follows:

I. Operational Factors

- A. Physical exertion to sweating (marching, lifting, etc.)
- B. Immobility (standing guard, riding in vehicle, sleeping)
- C. Vehicle operations (driving, fueling, operating with exhaust, helicopter wash, wind from vehicle movement).

II. Equipment Factors

- A. Clothing problems (not available, inadequate).
 - 1. Socks (wet, dirty)
 - 2. Boots (wrong type, defective, too tight)
 - 3. Other clothing problems
- B. Heaters (vehicle, tent)
- C. Other malfunctioning equipment

III. Human Factors

- A. Person factors: (lack of knowledge, failure to act, medical history)
- B. Group factor (wrong decisions made by coercion)
- C. Command/Leadership factor (e.g., orders not to use proper equipment)

IV. Accident Factors

- A. Break through or fall on ice
- B. Snow in clothing
- C. Water or fuel spilled.

Analysis of contributing factors was then implemented using the above listing.

RESULTS

Twenty-five case histories were obtained from the two surveys (16 from 1980 and 9 from 1981) and combined for analysis (Sampson, Stokes, Barr, Jobe and Hamlet, 1980; Sampson, Stokes and Stein, 1981). Table I summarizes the type and number of injuries of the combined cases. Sixty-eight percent of all cases involved frostbite (82% of these involved the feet, 18% other body areas), 20% were hypothermia and 12% were trench (immersion) foot. Notably, 68% of the injuries involved the feet.

A total of 80 specific elements were tallied across the 25 cases. The highest number of different elements for any one case was six, while the least was one element.

Table 2 gives a summary of the tally for each element and the general grouping factor. Forty-five percent of all tallied elements were classified as Operational Factors, 37.5% were Equipment Factors, 11.3% Human Factors and 6.3% were Accident Factors. The elements identified as having the highest frequency of occurrence were Immobility (23.7% of all elements), Physical Exertion to Sweating (20%), Vehicle Heater Failures (10%), wet socks (10%), and Personal Medical History (8.7%).

Element combinations were also examined for frequency of occurrence. Table 3 summarizes this analysis. The most frequent single element was Immobility (found in 76% of the 25 cases), the most frequent two element combination was Exertion and Immobility (48%), the three element combination was Heater Failure, Immobility and Personal Medical History (prior injury) (16%) and the four element combination was Heater Failure, Immobility, Improper Boots (too tight) and Personal Medical History (12%).

DISCUSSION

In reviewing the data as summarized, it becomes evident that weather conditions and operational activities conspire to produce the bulk of all injuries. The dominant sequence of events is troop activity where physical exertion leads to overheating and sweating which followed by immobility where soldiers hold their positions for several hours. Physical exertion results in wet clothing and socks which in turn leads to serious loss of heat. Review of cold injuries during the Korean Conflict also found these important etiological factors (Hanson and Goldman, 1959). For personnel operating vehicles, the greatest danger appears to be reliance on vehicle heaters which frequently malfunction. Armor vehicles then become heat sinks sapping most of the available heat from personnel who are not wearing appropriate clothing (combat boots instead of the vapor barrier boots) because it is difficult to operate vehicles wearing the arctic boot.

Prior medical history of cold injury or cold hypersensitivity was also a factor and this combined frequently with problems with boots. It is suggested that part of their problem is inadequate discipline in acquiring proper equipment.

Accidental factors played a smaller role than might be expected. There were

only a few identifiable incidences of injuries resulting from accidents.

Although Human Factors appear insignificant given the present analysis, it must be noted that a great deal of human decision-making is inherent in the other categories of assigned cause. Decisions to work personnel to physical levels producing sweating, then placing them in holding situations are involved in the incidence of cold injury. Human decision is also involved in maintenance of vehicle heaters and selecting boots. Thus, it is possible to attribute most system failures largely to human factors.

Given the above perspective, the current findings suggest that the best intervention might be to improve decision-making at the command and individual levels involving physical activity of soldiers and the maintenance and distribution of essential equipment.

REFERENCES

Hanson, H.E. and R.F. Goldman. Cold injury in man: a review of its etiology and discussion of its prediction. Military Medicine 1969, 134, 1307-1316.

Sampson, J.B., J.W. Stokes, J.G. Barr, J.B. Jobe, and M.P. Hamlet. Morbidity rates during a cold weather exercise. Technical Report HP-80, US Army Research Institute of Environmental Medicine, Natick, MA 01746, 1980.

Sampson, J.B., J.W. Stokes, and E. Stein. Injury and Illness during cold weather training. Technical Report, US Army Research Institute of Environmental Medicine, Natick, MA 01746 (in draft).

Table 3. Frequency of Factor Combinations.

<u>Number of Factors</u>	<u>Factors</u>	<u>Freq.</u>	<u>%(of Cases)</u>
One	Immobility	19	76.0
Two	Exertion + Immobility	12	48.0
Three	Immobility + Heater Failure + Person	4	16.0
Four	Immobility + Heater Failure + Person + Boots	3	12.0
		38	(of 25)

Table 1. Frequency of Cold Injuries.

<u>Injury</u>	<u>Freq.</u>	<u>%</u>
Frost bite	17	68.0
(Feet)	(14)	(56.0)
(Other)	(3)	(12.0)
Hypothermia	5	20.0
Trench (Immersion)		
Foot	3	12.0
Total	25	100.0

Table 2. Factors Associated with Cold Injuries.

<u>Factor</u>	<u>Freq.</u>	<u>%</u>
A. Operational Factors:		
1. Physical Exertion	16	20.0
2. Immobility	19	23.7
3. Vehicle Operation	<u>1</u>	<u>1.3</u>
	(36)	(45.0)
B. Equipment Factors:		
1. Clothing Problems		
a. Socks (wet/dirty)	8	10.0
b. Boots	6	7.5
c. Other	4	5.0
2. Sleeping Bag	4	5.0
3. Heater Failure	8	10.0
4. Other	<u>0</u>	<u>0.0</u>
	(30)	(37.5)
C. Human Factors:		
1. Personal (Individual)	7	8.7
2. Group/Team	0	0.0
3. Command/Leadership	<u>2</u>	<u>2.5</u>
	(9)	(11.3)
D. Accident Factors		
1. Ice/Water	2	2.5
2. Snow	3	3.7
3. Other	<u>0</u>	<u>0.0</u>
	(5)	(6.3)
E. Total:	<u>80</u>	<u>100.0</u>

Longitudinal Health Effects Associated with Aviation

Anne Hoiberg

Naval Health Research Center
San Diego, California 92138

During their flying time, aviators are exposed to many unique environmental conditions that can pose a threat to their well-being or health status. Nine of these environmental stressors include: (1) decompression and compression changes, (2) decreased oxygen in the breathing air, (3) acceleration, (4) heat stress, (5) radiation, (6) ozone, (7) noise, (8) vibration, and (9) circadian desynchronization. Several specific disorders or conditions that are associated with such environmental stressors are barotitis, hypoxia, atelectasis, dyspnea, hearing loss, and jet lag.

In addition to these health risks, another area of concern for aviators is that of the effects of aging. A controversy has persisted over the years that centers on the issue of whether or not to use chronological age as the primary criterion for determining the time for a pilot to retire. For civilian pilots, the calendar age of 60 has been designated as the mandatory time of retirement. In the Navy, the age of 45, until just recently, served as the basis for reclassifying qualified aviators from Service Group I to Service Group III, which stipulated that the reclassified pilot could not fly unless accompanied by a Service Group I aviator. At present, the factors associated with a Navy pilot's fitness are freedom from impairing disease, ability to perform, and the desire to continue flying. Although Navy aviators are highly selected and subjected to periodic stringent physical examinations, it seems likely that, because of the unique environmental stressors they encounter, their longitudinal health patterns would differ from other members of the aviation community and from other officers. The purpose of this study is to identify the health-related problems unique to aviators and the age of greatest risk for illness or injury. In other words, this study endeavors to compare the reasons for hospitalizations across three groups of the aviation community and to identify the physical and mental disorders associated with various ages of a pilot's life.

Method

Participants

Participants for this study were identified from data summarized from the Individual Flight Activity Reporting System file, which was made available to the Naval Health Research Center in San Diego by the Naval Safety Center in Norfolk, Virginia. This file contains information on the flights of all aviators who served on active duty for any period of time since 1968. Many of these individuals also have records that date back to the 1950s and early 1960s. For the purpose of this study, the following data were extracted from this file: total flight hours as a first pilot, co-pilot, or special crew; birthyear; designator, pilot or non-pilot status; and final fiscal year of flying.

Using these data on the file, the total aviation population ($N = 35,968$) was

divided into the following three groups: (1) pilots with at least one hour of flight time as either a first pilot or a co-pilot (n = 22,417); (2) officers who had at least one hour of flight time as a member of the crew (n = 9,483); and (3) all other members of the aviation community who had records of no flight hours as first pilot, co-pilot, or special crew member (n = 4,068).

Procedure

Records from these three groups were matched against the naval medical inpatient historical file maintained at the Naval Health Research Center. The data on this file, which contains all hospitalization records from July 1965 through December 1979, are obtained annually from the Navy Medical Data Services Center in Bethesda and are collapsed into chronological or new records for each hospitalized individual. For the purpose of this study, the following data were extracted from the medical inpatient file: diagnosis for each hospitalization (numeric codes were adapted from the Eighth Revision of the International Classification of Diseases or ICDA-8), date of hospital admission, and birthyear.

The three groups of the aviation community were then divided into 6-year age intervals using each individual's birthyear as the basis for the categorization. The numbers of hospitalizations recorded during the July 1965 through December 1979 time period were tallied for each specific numeric diagnosis, the 16 major diagnostic categories, and the overall total for the six birthyear groups within each of the three groups. In order to compute hospitalization rates for these separate groups, the populations at risk were determined by calculating the mean numbers of individuals on active duty across each group's specified time period. Annual rates per 10,000 were then computed for each group.

Results

Across the three aviation groups, there was a total of 7,465 hospitalizations recorded during the 14.5 years of this study, with 4,420 for the pilot sample, 1,840 for the special crew, and 1,205 for all other members of the aviation community. Included in these subtotals were the hospitalizations for 56 women which were excluded from subsequent analyses for this study.

Longitudinal Health Effects among Navy Pilots

In comparing hospitalization rates across the six age groups of pilots, the age group with the highest rates for nearly all categories was the youngest or those men who were born after 1953. These men had disproportionately high rates for the following diagnostic categories: Diseases of the Digestive System, Accidents, Poisonings, and Violence; and Diseases of the Musculoskeletal System and Connective Tissue. The reasons for the high rates of hospitalizations for digestive disorders, which was almost 5 times higher than any other category, were primarily because of problems associated with tooth development and eruption.

Other comparisons of rates across the six age intervals of pilots revealed small differences for almost all of the major diagnostic categories. The only exceptions were the elevated rates for Diseases of the Digestive System for the 1947 to 1953 age group (although it should be emphasized that the rate was 11 times lower than that for the youngest group) and the only slightly higher rate for Diseases of the Circulatory System among aviators born prior to 1926.

Longitudinal Health Effects among Special Crew Members and Others

By way of contrast, the hospitalization rates across the six age intervals of special crew members were considerably more variable, with the highest overall rates observed for the oldest group (birthyear prior to 1926) and the youngest groups with a birthyear subsequent to 1953. Similar to the pilot population, the youngest group of the special crew category had elevated rates for Diseases of the Digestive System and Accidents, Poisonings, and Violence. The highest rates for the oldest group were for Mental Disorders, Diseases of the Digestive System, Diseases of the Musculoskeletal System, and the accidents category. Differences in rates across the other four groups, or those born between 1926 and 1953, showed only minimal variability for each of the 16 diagnostic categories.

In comparing hospitalization rates across age intervals for those members of the aviation community who had records of no flight hours, the rankings were quite similar to the other two groups in that for most of the age intervals the highest rates were observed for Diseases of the Digestive System and Accidents, Poisonings, and Violence. Hospitalizations for the category of digestive diseases, however, tended to be primarily for hernias which contrasted with the reasons of tooth development and eruption for the pilot group.

The overall comparisons across the three groups showed that pilots were the healthiest members of the aviation community whereas intragroup comparisons revealed that the youngest subgroup of pilots was at greatest risk for accidental injury and disorders of the teeth. The health effects of aging, as reflected by increased hospitalization rates for Diseases of the Circulatory System and Diseases of the Musculoskeletal System, were not evidenced by elevated rates for the older pilots but were observed among the special crew members.

Disorders Associated with Environmental Stressors

In the other phase of this study, the objective was to determine whether or not pilots and special crew members had higher hospitalization rates than the "others" group for the diseases and disorders identified as related to the environmental stressors of flying. For these comparisons, hospitalization rates were computed for each of the three groups for such disorders as pneumothorax, alveolitis, hearing loss, cataracts, other respiratory or lung conditions, and joint pain. Results showed that there were no significant differences among the three groups for and of the specified disorders.

Discussion

Results of this study indicated that Navy first pilots and co-pilots of any age had the lowest hospitalization rates within the active duty aviation community for most physical and mental disorders. The youngest pilot group, however, did have relatively high rates for the accidents category and digestive disorders associated with tooth development and eruption. In comparing rates across the three groups of men in just the oldest age interval, the pilots had the lowest rates for all diagnostic categories which also were not significantly higher than those for pilots in any of the other age intervals. Thus, the findings of this study provided support for results reported elsewhere (e.g., the 1,000 aviator research project) that pointed up the excellent health status of pilots. Such results were not surprising in view of the fact that pilots must meet high standards of selection and medical surveillance throughout their careers.

Other results revealed that the specific disorders associated with flying did not differentiate one group of the aviation community from the other two groups. In comparisons with other Navy personnel, however, the rates for hearing loss and hearing problems were considerably higher whereas the other rates tended to be fairly comparable. At present, the results of this study are being compared with the medical inpatient data for all other officers to determine whether or not the hospitalization rates for the three groups of the aviation community differ significantly from all other officer groups in the Navy. This ongoing study will provide the needed data to identify those health problems associated with all other officer designators. Results of this project will furnish the Medical Department of the Navy with information needed to form the basis for the development of health care delivery and intervention programs.

MAINTENANCE PERFORMANCE RESEARCH IN THE ARMY

John F. Hayes

US Army Research Institute

The U. S. Army Research Institute (ARI) has a long range program of research underway directed at the improvement of job performance of maintenance personnel. Figure 1 represents what are considered to be the major factors affecting that performance. These are: institutional training or that initial formal training received in a school setting prior to the individuals first assignment; unit training, referring to the training conducted in the unit to refresh existing skills and impart new ones; management, meaning the provision and control of resources and work flow; the equipment that has to be maintained, both its condition and complexity; and the quality and extent of job aids provided to the technician. This paper will describe a current project that focuses on the unit training and management aspects of maintenance performance.

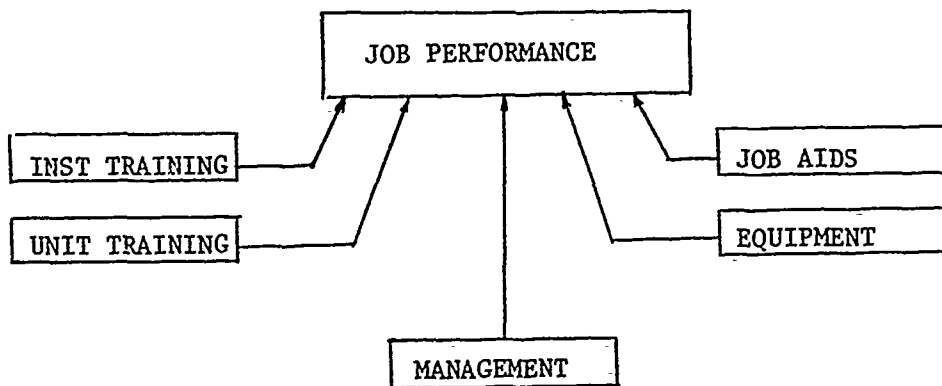


Fig. 1. Job Performance Elements

Technical training has been and continues to be difficult for units to deliver. There are three major obstacles present that account for this in maintenance units: the mission, resources, and personnel shifts. The mission of maintenance units requires that first priority be given to keeping equipment operational and this is generally a full-time commitment in itself. Resource constraints exist in terms of a lack of qualified personnel to conduct technical training and often in terms of a lack of equipment and training materials as well.

Finally, individual technical training presents a management challenge in trying to keep abreast of the myriad of constantly changing individual requirements, due to personnel turnover and shifting needs as personnel gain experience.

The ultimate impact of these constraints on maintenance performance can be plainly seen in some of the results from a recent ARI project that examined how maintenance personnel sought and used technical information. Detailed observations were made of some 300 vehicle repair tasks being performed by military mechanics. In addition to recording their information usage, the work of the mechanics was evaluated in terms of errors committed in the performance of the task. One type of measure employed was whether the mechanic left any errors in his work upon completion that were judged to be serious by the trained observer. Figure 2 shows the results of these observations for two major categories of tasks, those requiring the use of special tools or technical specifications (ST/TS) and those not requiring ST/TS. As can be seen, the error rate for the more complex tasks would be unacceptable by any standards.

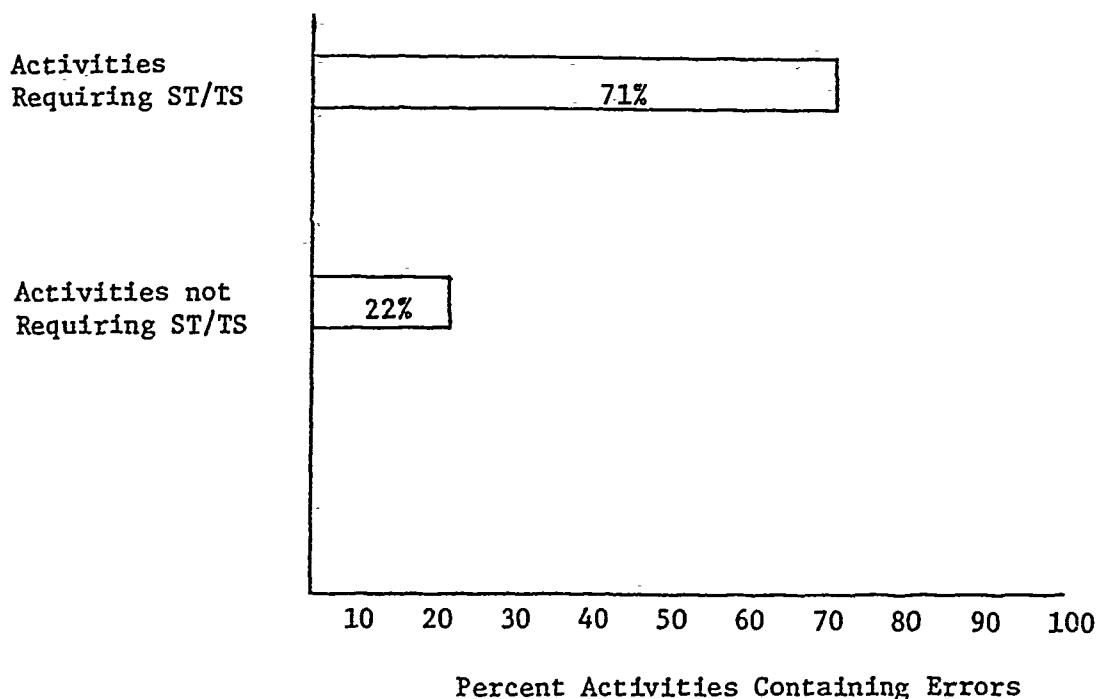


Fig. 2. Error Performance by Type Activity

A further analysis of these data examining the impact of job experience on performance was even more revealing. Prior to performing a task being observed, the mechanic was asked to estimate how many times he had previously performed that task. Figure 3 shows the results obtained for checkout activities,

indicating that about two-thirds of the time neither experienced nor inexperienced personnel properly checkout their work upon completion.

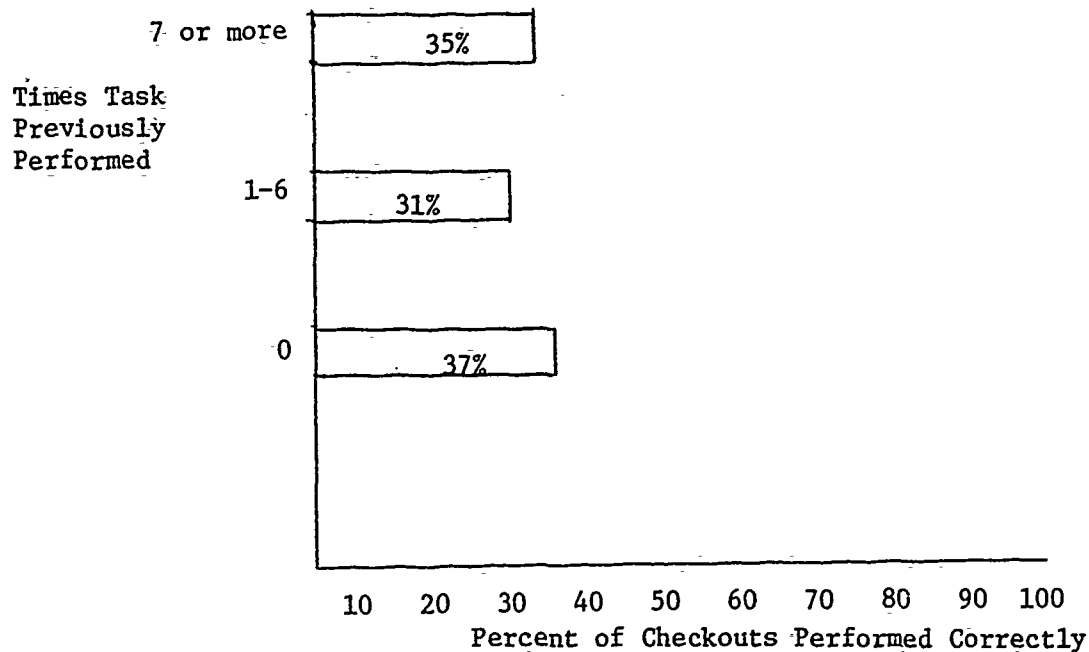


Fig. 3. Checkout Performance by Task Experience

Further, Figure 4 indicates that not only does the quality of work not improve with experience, it tends to decline. What these data reflect, of course, is that without training and feedback on performance the individual has no basis or reason to change his performance.

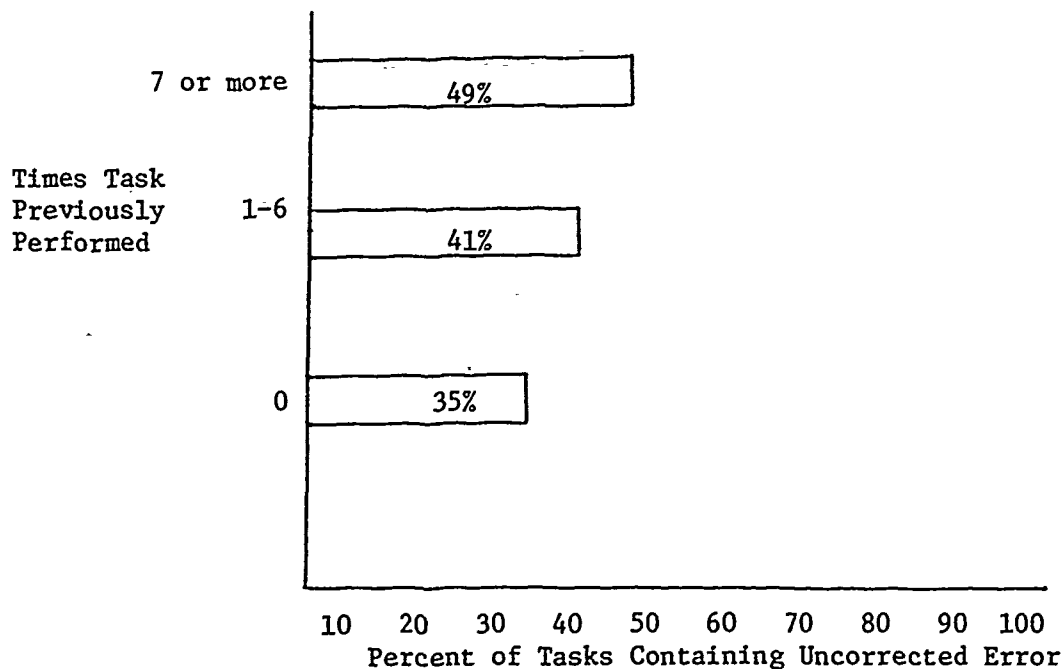


Fig. 4. Errors by Task Experience

The situation portrayed by these data will require corrective actions in a number of areas to produce significant changes in performance. The balance of this paper describes one such effort that is being conducted by ARI, which is called the Maintenance Performance System (MPS). This system is intended to provide a basic component necessary for the development of solutions to unit maintenance training problems, which is information. The MPS introduces a small computer into the unit so that the thus far overwhelming task of tracking and managing individual training requirements can be readily accomplished. Manual systems tried by the Army for this purpose have proven largely unsuccessful due to their administrative and bookkeeping load, thus there is no solid information on which to base training when the opportunity does arise.

MPS continuously tracks two types of information on each maintenance technician in the unit: formal training and testing accomplished and qualifying job experiences. These individual accomplishments are then aggregated and compared with a master list of unit skill requirements so that two basic types of information are available at all times. These are what any specific individual needs in terms of training or experience and what the unit needs based upon the available skills of the current mechanic population.

The system operates as follows. It is first loaded with a listing of all the maintenance tasks, by equipment and Military Occupational Specialty (MOS), that a particular type of unit must perform. Then a roster of unit technical personnel is added. This, plus the program disk, are the enabling components of the system. As work is performed in the shop, the names of each mechanic performing each task and the number of hours spent are recorded by the supervisor. This type of job performance data represents new information for Army maintenance, as the emphasis in such systems to date has been to track equipment rather than personnel. In addition to this information, the supervisor must also indicate whether the work of each individual was of sufficient substance and quality that it represented a positive skill development or sustainment experience. This then provides continuous information on individual job experience. The supervisor must also provide information to the system on individual technical training and testing experiences as they occur, and provide an assessment of the training value obtained. Simple input forms are provided for this purpose.

The recording and evaluating of individual job performance and training represents the bulk of the additional work required by line personnel by MPS. The balance of the work is performed by an administrative clerk assigned to operate the system for a battalion. The MPS clerk picks up the data cards from the maintenance shops on a daily basis and updates the files for each company. With this information, the MPS programs can provide variety of information on individual and unit training requirements tailored to the particular unit. The MPS clerk has a schedule for producing and delivering these reports to battalion personnel, from the commander to the individual technicians. The major types of reports provided are:

- training requirements summaries
- skill and growth indexes
- individual skill histories
- man-hour availability and use
- average direct man-hours per job

The training requirements summaries are the principal output for the purpose of managing technical training. This report lists the prioritized individual training requirements by task, based on a comparison of individual skill capabilities with the master list of unit requirements. It lists the individuals requiring training on each task and provides references to the source of relevant training programs and materials available to the unit.

The skill and growth summaries provide aggregate reports to supervisors and commanders that show the status and trends in skill development for the unit. The individual skill histories are provided to the individual technicians so that they can see what progress they are or are not making in becoming fully proficient in their jobs. The other reports identified above provide additional management information that permit evaluation of personnel utilization efficiency and the development of unit-based time standards for task performance. This direct and immediate feedback to all levels of unit personnel is an important feature of MPS and is considered to greatly enhance the quality of the input data provided. Supervisors know that what they provide to the system will affect the utility of the output they will receive in return.

The MPS has been under development and in operation in one Army direct support maintenance battalion for over a year and in operation in a second for several months. Experience gained to date indicates that cost of operating the system is quite acceptable to the units and the outputs have been judged to be of considerable value. The most significant use of the MPS to date has been to guide job assignments, especially for high density MOS's. With the information MPS provides, the supervisor can readily identify who needs specific types of job experience, so that when the mission requirements permit, he can properly use the job environment for training and development. MPS generated information is also being used to guide off-line individual training, although the opportunities for this type of training remain limited due to the demands of the mission and availability of personnel. MPS data have made this latter problem more explicit by providing objective data indicating that on the average mechanics are available for maintenance job performance less than 50% of the time, due to outside demands that the Army imposes upon them.

One of the most important outcomes sought for MPS is to help break the pattern of early incorrect task performance becoming internalized by the technician due to lack of supervision and feedback on performance. It attempts to do this by identifying training requirements, highlighting early performance trials, and requiring overt task evaluation by the supervisor of each performance of his personnel. As we proceed with MPS development and also its extension into the organizational level of maintenance we will be seeking evidence of whether this is happening. The task that we are addressing at present is how to routinely capture more detailed job performance data that will permit such an evaluation.

On a larger scale, MPS is seen as the initial component in a total training system for the future. Current efforts are being directed at extending it into computer-based training delivery, to assist in overcoming the training resource problems faced by units. In addition, the data base represented by MPS is known to have considerable value as reference data for systems design efforts. Its availability to equipment, personnel, and training system designers of the future is considered essential. While this need has been recognized for some time,

technology is now becoming available to make it feasible to automate its acquisition and dissemination to a much greater extent than previously possible. This is required since a major impediment to the success of such information systems in the past has been the extent to which they have been dependent on manual inputs and processing. As each of the steps in the process become automated, we come nearer to being able to develop the large scale integrated training and performance information systems that are ultimately needed.

FIREFIGHTING PERFORMANCE OF
NAVY MALE AND FEMALE SHIPBOARD CREWMEMBERS

ROSS L. PEPPER AND MARK D. PHILLIPS
NAVAL OCEAN SYSTEMS CENTER, KAILUA, HAWAII

ABSTRACT

A study of fire fighting performance was conducted as part of an on-going effort to identify potential gender-based human engineering problems resulting from the introduction of women to shipboard assignments in the Navy. Results of experimental simulations of fire fighting with a CO₂ extinguisher and starting a P-250 pump were consistent with earlier questionnaire data which found female difficulties with tasks involving upper torso strength and grip strength. In the experiment, females registered significantly lower force values on the P-250 starter pull than males; with lighter females exerting less force than heavier females. Females also exhibited a different pulling strategy than males. Unlike females, weight was unrelated to the pulling performance of males. The results revealed that 85% of the females and 5% of the males sampled would be incapable of starting the P-250 pump. Further work focusing on gender problems involving ship motion is recommended.

FIREFIGHTING PERFORMANCE OF NAVY MALE AND FEMALE SHIPBOARD CREWMEMBERS

ROSS L. PEPPER AND MARK D. PHILLIPS

INTRODUCTION

The 1970's brought many demographic shifts in the make-up of the work force in the United States. During the past decade, women have entered diverse and non-conventional careers as construction workers, law enforcement officers, and heavy machine operators. The trend toward expanded female work roles is likely to continue into the 1980's.

This demographic shift in the composition of the civilian work force has its parallel in the military sector. While increasing cultural acceptance has played a significant role in increased female enlistment, there have also been practical reasons. Since the mid 1970's, the number of military-aged males in the United States has been consistently declining. This population shift has been accompanied by the abolition of the draft in 1973. These two factors have forced the armed services to re-examine the utilization of women in non-traditional areas.

The role of women in the armed forces has expanded from traditional clerical and nursing positions to include almost all non-combatant military jobs. The Navy has, in turn, expanded its assignment of women to include permanent shipboard duty assignments on auxilliary support ships, such as tenders, repair ships, tugs, etc.

Recent instruction from the Secretary of the Navy (Secretary of the Navy, 1979), stated that, "It is the policy of the Department of the Navy that women members, officers and enlisted, will be assigned to billets commensurate with their capabilities to the maximum extent practicable." The only restriction placed on female personnel utilization is that women may not be permanently assigned to a combatant vessel. Women may serve temporary duty on any ship in the Navy, provided that it is not expected to have a combat mission during the period of temporary duty. This expansion of the role of women aboard ship has led to female personnel being assigned to over twenty Naval ships to date.

The introduction of women into a workplace designed for use by males has been shown to cause work performance difficulties directly related to the deficiencies in workplace layout (Ketcham-Wiedle and Bittner, 1977). The influx of women into shipboard billets therefore raises questions about the adequacy of shipboard equipment design for female use. Ship fittings and equipment were designed to accommodate the male user. No consideration was given to human engineering standards that would accommodate both males and females in ship systems.

Background

This study is part of an on going program aimed at determining the extent of human engineering problems which may exist as a result of assigning women aboard ships. The problem identification phase of this program employed questionnaires, interviews, and observational techniques to identify and describe shipboard human engineering problems due to the anthropometric and biomechanical differences of the sexes (Pepper and Phillips, 1980).

That descriptive effort revealed that females currently on board lack experience using many ship fittings, equipment, and protective clothing. Very few of the women

had extensive sea duty. Since the males were often selected or volunteered to handle the more physically-demanding jobs, the resulting lack of exposure of women to many ship system elements potentially reduced the reported incidence of equipment-related problems for the female complement. Additionally, the high motivational level of the female sailors might have masked some equipment problems in the self-report indices.

Even with these limitations of the data, several items of ship equipment, clothing, and fittings emerged as being deficient for use by females. Protective gear such as the oxygen breathing apparatus (OBA), safety harness, life preserver, and foul weather gear do not adequately fit many of the women on board. Hand tools and fire fighting equipment, such as the P-250 pump, were found to be difficult for many of the women to operate. Ship fittings such as escape scuttles and water tight doors were also reported as problematical. Factors found to be associated with these gender based human engineering difficulties included the disparities in grip and upper torso strength and reach envelope between the sexes. These results are described in greater detail elsewhere (Pepper and Phillips, 1980).

In an effort to verify by objective means, problem areas identified in the qualitative data, an empirical investigation of female and male performance in selected shipboard tasks was designed. It was felt that an experimental approach could circumvent the biases inherent in the qualitative data due to differences in experience and motivation found between the sexes. Additionally, an experimental analysis is necessary to bridge the potential gap between perceived and actual human engineering related difficulties in the shipboard environment.

The experiment was carried out at the Fire Fighting (FF) school located at the U.S. Navy Fleet Training Center in San Diego. Fire Fighting school provides an opportunity to examine critical, real world shipboard tasks in a somewhat controlled environment. At FF school, extremely realistic simulations of shipboard fires are conducted. The same equipment, tools and protective gear issued to fight fires on board ship are used in these simulations. Two tasks were chosen for evaluation at FF school; these were fighting a class C (electrical) fire with a standard Navy CO₂ portable extinguisher and starting a P-250 pump.

EXPERIMENT 1 - CO₂ TASK

Method

Subjects. Subjects were 16 male and 16 female enlisted Navy personnel. The female sample ranged from the 5th to the 70th percentile of military females in weight (E. Churchill, T. Churchill, McConville, and White, 1977). This range was 47.6 kg to 63.5 kg with a mean weight of 54.3 kg. The male sample ranged from the 10th to the 95th percentile in weight (McConville, E. Churchill, T. Churchill, and White, 1977). The range was 59.0 kg to 91.6 kg with a mean weight of 76.7 kg. No measures of relative experience with fire fighting equipment were obtained. The experiment was part of an on going fire fighting and damage control training program in which all of the subjects were required to participate.

Procedure.

Subjects were randomly assigned to one of 32 positions in a line-up inside a bunker which was physically identical to a lower port side ship compartment. As each subject came to the front of the line, they were required to observe and verbally report a simulated class C fire. The fire was actually fueled by one quart of diesel fuel splashed over an electric motor frame. After the subject made the report, he or she carried a 27.7 kg CO₂ extinguisher a short distance (2.1 - 2.4 m) and fought the blaze, using techniques which had been perviously demonstrated. All subjects wore foul weather gear including boots and gloves for this task.

A time measure was taken with a stopwatch from the time the CO₂ extinguisher was activated until the fire was reported to be out. As each subject completed the task, they reported their weight to the experimenter. The experimenter then recorded this weight, the time required to extinguish the fire, and noted whether the subject was a male or a female.

Result and Discussion.

In order to view the effects of size as well as sex on task performance, males

and females were grouped into two weight categories. While grouping and matching according to percentile ranks was desired, limited subject availability only allowed grouping according to a median split along the obtained weight distribution. While grouping in this manner is somewhat arbitrary, it does provide a limited means of investigating the very important size variable in these tasks within the real-world constraints of this study. The female sample was split at the median value of 54.4 kg, the male sample was split at 74.8 kg.

Grouping the data in this manner showed that females did tend to take longer than males to put the fires out, and that lighter (Group 1) females took longer than heavier (Group 2) females. The mean times (in seconds) for the CO₂ task are shown in Table 1.

Table 1.
Time to Extinguish Fires With CO₂ Bottle

Gender	Group 1		Group 2	
	Mean (s)	SD	Mean (s)	SD
Females	18.3	6.81	14.50	2.93
Males	13.00	5.04	12.25	5.70

A 2x2 (sex x weight) ANOVA performed on these data show that the main effect of sex approached statistical significance, $F(1, 28) = 3.85$, $p < 0.1$. The effects of weight group and the interaction of weight group and the interaction of weight group by sex did not approach significance.

While size matching was not possible in this study (as noted earlier), there was some degree of overlap in the obtained weight distribution of females and males. Examining the performance of the subjects listed in Table 2, it can be seen that weight matched females also took longer than their male counterparts to extinguish the fires.

Table 2.
Times to Extinguish Fires for Weight Matched
Male and Female Subjects with CO₂ Bottle

Female		Male	
Weight (kg)	Seconds	Weight	Seconds
60.0	11	60.0	8
63.5	15	63.5	8
63.5	19	65.8	16

It should be noted that many females were observed to have great difficulty carrying the extinguisher out of the bunker (especially lifting it over a 0.3 m bottom door lip) after the fire fighting task was completed. This difficulty was not apparent in the male subjects. Due to constraints imposed by the schedule of the overall training program it was not possible to add a lifting and distance carrying component to the fire fighting task. If this additional component was added, the results obtained for the CO₂ task may have revealed much greater sex based differences than those obtained.

EXPERIMENT 2 - P-250 TASK

Method.

Subjects. Subjects were 20 male and 20 female personnel. The females ranged

from the 5th to 90th percentile of military females in weight (47.6 kg to 70.3 kg) with a mean weight of 55.3 kg. The males ranged from the 15th to 95th percentile (61.2 kg to 91.6 kg) with a mean weight of 76.3 kg. (Percentile ranks were derived from Churchill et al., 1977 and McConville et al., 1977). As in the prior task, all subjects were required to participate in the study as part of an on going training program. Most of the P-250 subjects also participated in the CO₂ task.

Procedure. The P-250 task involved pulling a starting cord on a P-250 pump. The P-250 is a portable, gasoline powered pump used for dewatering flooded compartments and fire fighting. The pump is contained in a steel frame which measures 0.76 m x 0.91 m x 0.91 m. Without the gas tank, the pump weighs 66.7 kg. It is started by pulling a cord which is similar to a starting cord found on an outboard engine. Standard Navy issued work clothes were worn during this task. No work gloves were used.

A Dillon load cell (#0.200) wired to a Dillon load cell readout Model LAB 4598) was attached to the starter cord of the pump. This apparatus permitted a direct reading of the pounds of force exerted on the starter cord. The fuel line was detached from the pump engine so as to prevent ignition during the experiment.

Subjects were chosen in random order. They were directed to pull on the starter cord in a sincere attempt to start the pump. Subjects were instructed to use any stance or grip that felt comfortable to them. All subjects were allowed three attempts; the best of the three tries was recorded. After completion of the task, each subject reported their weight to the experimenter. The experimenter recorded the force level achieved, the subject's weight, and noted the subject's sex.

Results and Discussion.

A median split as described previously was used to group the P-250 task sample. The female and male samples were again split at their respective median values of 54.4 kg and 74.8 kg.

Since force values were given by the experimental apparatus in pounds, they were converted to newtons, their SI equivalent. Females registered lower force values than males on this task, with the Group 1 females registering lower readings than the Group 2 females. The mean force values registered are shown in Table 3. To put these means in perspective, baseline measures on this task indicated that a well maintained P-250 requires 71-89 N to start. The data obtained indicate that 85% of the women and 5% of the men sampled would not be able to start the P-250 even under optimal conditions.

Table 3.
Pulling Force Exerted in the P-250 Task

Gender	Group 1		Group 2	
	Mean (N)	SD	Mean (N)	SD
Females	46.26	14.23	62.72	13.97
Males	88.07	8.32	86.29	12.45

A 2 x 2 ANOVA conducted on these data indicate that the effect of sex of the subject on task performance was significant, $F(1, 36) = 68.64, p < 0.001$. The effect of sex accounted for almost two-thirds of the variance in this data set as is shown by $\eta^2 = 0.61$. The sex by weight group interaction was also significant, $F(1, 36) = 5.34, p < 0.05$. In this case, the strength of the interaction was $\eta^2 = 0.05$. The main effect of weight group was not significant.

The results supported the expected outcome that males could exert more starting force than females. Additionally, while weight makes no difference in force exerted by males, it does impact the performance of females. Further evidence of this effect can be gained by examining the correlations between weight and performance in the male and female samples. The Pearson product moment correlation obtained for the female sample showed a significant positive correlation between subjects' weight and force exerted, $r = 0.51, p < 0.05$. The correlation obtained for the male sample did not approach significance, $r = 0.01$.

The sex by weight group interaction may result in part from the different strategies males and females employed during the P-250 task. Males tended to exert a single explosive motion while females applied a more continuous pull on the cord. The larger overall mass of the Group 2 females compared to the Group 1 females may in some degree compensate for the lower upper torso strength females have than males (Laubach, 1976). The greater mass can thus be converted into increased force when using the continuous force application technique typical of females. The greater male upper torso strength, exhibited by use of the explosive motion, may have negated the effects of weight group (mass) on male performance.

The performance of the weight matched subset of subjects was consistent with the main effect of sex in the overall results. As Table 4 indicates, females exerted less force than did the weight matched males.

Table 4.
Force Exerted by Weight Matched Female and
Male Subjects in the P-250 Task

Female		Male	
Weight (kg)	Force (N)	Weight (kg)	Force (N)
62.6	40.03	61.2	84.51
63.5	75.62	63.5	80.06
63.5	53.38	65.8	102.30
70.3	62.27	70.3	88.96

CONCLUSION

The empirical investigation reported here adds substantive support to the results of the Pepper and Phillips (1980) problem description study. For example, females (especially Group 1 females) registered lower force values in starting the P-250 pump and exhibited a different pulling strategy than males did. This finding specifically supported earlier questionnaire reports of female difficulty in starting the P-250. These findings also lend support to questionnaire reports of female difficulty with similar tasks involving a combination of upper torso strength and grip strength.

Also apparent in these two data sets is that the population least accommodated by current clothing, equipment, and ship fitting configurations is the Group 1 (small) females. Here again, earlier qualitative reports of difficulties by this sub-set of subjects was supported by the experimental results.

The empirical methods used here allowed a systematic identification of equipment problems that would not be possible using paper and pencil methods which attempt to match human performance data with equipment use requirements (e.g. Palla, Jones, Lovett and Porter, 1980; Advanced Marine Enterprises, 1980). For example, females used a different pulling strategy than males during the evaluation. This may have training as well as design implications which would not be apparent from data which simply show male/female force production capabilities. It should also be noted that biomechanical data (e.g. Laubach, 1976) is task and position specific. In other words, the values obtained will often not generalize to real world conditions where operators have limits on their body position, grip surface, etc.

While the results reported here are consistent, they must be viewed as exploratory and incomplete due to the methodological limitations imposed by the experimental setting, and the absence of the potentially critical performance variable of ship motion. The well documented anthropometric differences of the sexes (Kennedy, 1978; Robinette, Churchill and McConville, 1979; White, 1979) suggest that there may be gender specific responses to ship motion. More work is clearly warranted in this area to further elucidate these differences, and to translate them into design guidelines where applicable.

REFERENCES

- Advanced Marine Enterprises. Suitability of Hull and Damage Control Equipment and Systems for Women Aboard Ship. Arlington, VA.: March, 1980.
- Churchill, E., Churchill, T., McConville, J.T., and White, R.M. Anthropometry of Women of the U.S. Army, Report No. 2, Basic Univariate Statistics. Natick, Mass.: U.S. Army Natick Research and Development Command, Technical Report TR-77-024, 1977
- Ketcham-Wiedl, M.A. and Bittner, A.C., Jr. Anthropometric Accommodation of a Female Population in a Workplace Designed to Male Standards. Point Mugu, CA: Pacific Missile Test Center, Technical Report TP-76-3, 1976.
- Laubach, L.L. Muscular Strength of Women and Men: A Comparative Study. Wright Patterson Air Force Base, Ohio: Aerospace Medical Research Laboratory, Technical Report AMRL-TR-75-32, 1976.
- McConville, J.T., Churchill, E., Churchill, T., and White, R. Anthropometry of Women of the U.S. Army - 1977. Report No. 5, Comparative Data for U.S. Army Men. Natick, Mass.: U.S. Army Natick Research and Development Command, Technical Report TR-77-029, 1977.
- Palla, R. L., Jr., Jones, C.E., Lovett, C.D., and Porter, L.G. Evaluation of Selected Navy Equipment for the Women Aboard Ship Program. Washington, D.C.: National Bureau of Standards, Technical Report NBSIR-79-1949 (Navy), January, 1980.
- Pepper, R.L. and Phillips, M.D. Naval Architectural Research for Women Aboard Ship. San Diego, CA: Naval Ocean Systems Center, Technical Report NOSC-TR-658, March, 1981.
- Secretary of the Navy. Assignment of Women in the Department of the Navy. Washington, D.C.: Instruction SECNAVINST-1300.12, April, 1979.
- White, R. M. The anthropometry of United States Army men and women: 1946-1977. Human Factors, 1979, 21(4), 473-482

THE DEVELOPMENT OF THE AIR FORCE CIVILLIAN JOB PERFORMANCE APPRAISAL SYSTEM

John A. Guerrieri
Air Command and Staff College
Maxwell AFB, AL 36112

Abstract

The purpose of this paper is to describe the development of the Air Force Job Performance Appraisal System. It provides 1) a cursory discussion of appraisal literature; 2) a description of the constraints placed on the system design by the Request for Personnel Research, the Civil Service Reform Act of 1978, the Office of Personnel Management, and 3) the results of the preimplementation tests and other implementation considerations.

Background

In response to a Request for Personnel Research (RPR 76-40) entitled Supervisory Appraisal System for Air Force Employees, the Air Force Human Resources Laboratory (AFHRL) initiated research to develop a system for objectively evaluating the job performance of Air Force civilian employees. In October 1978, Public Law 95-454, The Civil Service Reform Act of 1978 (CSRA), was passed. The CSRA mandated the development of a new civilian appraisal system for Federal employees. The new law provided the regulatory guidelines and specific requirements which must be followed in the development of appraisal systems for government agencies. To meet the requirement for an appraisal system for all civilian employees, AFHRL developed three subsystems: 1) The Senior Executive Appraisal System (SEAS), 2) The General Manager Appraisal System (GMAS), and 3) The Job Performance Appraisal System (JPAS). This paper discusses the design and development of the Job Performance Appraisal System for all General Schedule (GS) and Federal Wage System (FWS) employees in grades 1-15 and GS employees in grades higher than 15 not covered by SEAS or GMAS. The other two systems are reported in detail by Guerrieri (1981) and Cowan, Thompson, Guerrieri, and Vitola (1981).

Literature Review

The first phase in the development of the Air Force Job Performance Appraisal System was a comprehensive review of the appraisal literature. To facilitate a discussion of this review, this paper arbitrarily groups the various appraisal systems into three functional categories: 1) perceived behavior, 2) predetermined goals/standards, and 3) traits assessment. Although these categories are not all encompassing, they do provide a convenient rubric for delineating the sundry appraisal systems.

THE DEVELOPMENT OF THE AIR FORCE CIVILLIAN JOB PERFORMANCE APPRAISAL SYSTEM

John A. Guerrieri
Air Command and Staff College
Maxwell AFB, AL 36112

Abstract

The purpose of this paper is to describe the development of the Air Force Job Performance Appraisal System. It provides 1) a cursory discussion of appraisal literature; 2) a description of the constraints placed on the system design by the Request for Personnel Research, the Civil Service Reform Act of 1978, the Office of Personnel Management, and 3) the results of the preimplementation tests and other implementation considerations.

Background

In response to a Request for Personnel Research (RPR 76-40) entitled Supervisory Appraisal System for Air Force Employees, the Air Force Human Resources Laboratory (AFHRL) initiated research to develop a system for objectively evaluating the job performance of Air Force civilian employees. In October 1978, Public Law 95-454, The Civil Service Reform Act of 1978 (CSRA), was passed. The CSRA mandated the development of a new civilian appraisal system for Federal employees. The new law provided the regulatory guidelines and specific requirements which must be followed in the development of appraisal systems for government agencies. To meet the requirement for an appraisal system for all civilian employees, AFHRL developed three subsystems: 1) The Senior Executive Appraisal System (SEAS), 2) The General Manager Appraisal System (GMAS), and 3) The Job Performance Appraisal System (JPAS). This paper discusses the design and development of the Job Performance Appraisal System for all General Schedule (GS) and Federal Wage System (FWS) employees in grades 1-15 and GS employees in grades higher than 15 not covered by SEAS or GMAS. The other two systems are reported in detail by Guerrieri (1981) and Cowan, Thompson, Guerrieri, and Vitola (1981).

Literature Review

The first phase in the development of the Air Force Job Performance Appraisal System was a comprehensive review of the appraisal literature. To facilitate a discussion of this review, this paper arbitrarily groups the various appraisal systems into three functional categories: 1) perceived behavior, 2) predetermined goals/standards, and 3) traits assessment. Although these categories are not all encompassing, they do provide a convenient rubric for delineating the sundry appraisal systems.

One current approach to performance appraisal posits that job performance can be accurately determined by comparing observed or perceived behavior to preselected behavioral statements which indicate various levels of achievement. For example, at the end of the appraisal period, the supervisor of a bookkeeper would select one of the following statements as an indicator of typical employee behavior: 1) Is the sort of employee who would balance work sheets on the first try, 2) Is the sort of employee who would commonly make substantial arithmetic errors, 3) Is the sort of employee who would recalculate figures using various methods to ensure validity. Numerous variations of this format exist. Some systems provide up to seven examples of behavior for each task. Smith and Kendall (1963) referred to these statements as Behaviorally Anchored Rating Scales (BARS). Appraisal systems which fall within this category usually require an indepth job analysis of each of the jobs to be evaluated. Once the analysis is completed, the system designers specify behaviors or critical incidents which depict various levels of job performance. Although this system is considered to be objective and effective, the amount of time required to thoroughly develop the required behavioral anchors for each task of each job is often prohibitive.

A second approach utilizes predetermined objectives or standards as the criterion for measuring performance. Performance in this case is measured as a function of production or some other measure of organizational achievement - behavior is unimportant. This approach to evaluating performance is usually associated with Management By Objectives (MBO) systems, (Odiorne, 1965). This approach also stresses the need for some type of job analysis in order to delineate realistic objectives or standards the employee expects to meet. At the end of the rating period, the supervisor assesses the employees performance in relation to the predetermined plan, thereby determining if the employee met, did not meet, or exceeded his/her goals or standards. Although this approach also requires some type of job analysis, the initial startup time requirement is substantially less than that of the previous system because the analysis is usually performed by the supervisor and employee while developing their plan.

The third approach involves the rating of various traits such as creativity, leadership, attitude, enthusiasm, determination, etc. At the end of the appraisal period, the supervisor subjectively assesses the amount of a given trait exhibited by the employee during the rating period. This type of appraisal is especially useful when normative comparisons are required. It is also the type of system which is usually preferred by supervisors because it affords them the most flexibility in determining the employee's rating. However, it is often perceived by the employee as lacking objectivity.

that 1) the system provided a viable method for appraising the job performance of Air Force civilians, 2) there was no systematic discrimination against women and minorities inherent in the system, 3) the system effectively measured actual job performance, 4) the participants perceived the system as more objective than the current appraisal system, 5) the system design aided in minimizing rater inflation, and 6) the job related communication between supervisor and employee was enhanced.

The field test also surfaced some problems: 1) the participants perceived the process of developing objective work plans as too time consuming; 2) many indicated it was not possible to write objective standards for their jobs; 3) the training did not sufficiently prepare them for developing an effective work plan; and 4) with the standards written at only one level, the supervisors had difficulty determining whether an employee exceeded the standard or far exceeded it. For an indepth review of these studies, see Watson, Guerrieri, and Earles (1980).

Implementation Considerations

In order to ensure effective implementation and utilization of the JPAS, a number of requirements must be considered. Perhaps the most important of these is training. Without a thorough understanding of the system, Air Force employees will be unable to participate effectively in the appraisal process. If they do not completely understand the system and feel their efforts are ineffective, they will perceive the system as unworkable and inappropriate.

Training is the link in the system that makes it run. The system can only be as strong as the people who operate it and those people learn the proper procedures only through training and practice. It is imperative that all employees be trained and that new employees are trained as they enter the system. To ensure currency and reliability of the appraisal system, training should be periodically reviewed and updated as necessary. Participants should also receive periodic refresher training.

Another consideration crucial to efficient implementation is the necessity for high-level support. Without the support of Headquarters USAF, and the Major Command, Separate Operating Agency, and Direct Reporting Unit Commanders, the individuals affected by the system will perceive it as relatively unimportant. The supervisors and employees would believe that the system was not worth the time required to use it effectively; or that the new system was something that would go away, could be ignored, or changed at will.

Conclusion

The JPAS was developed after a thorough review of the appraisal literature and comprehensive field testing in response to a Request for Personnel Research submitted by the Directorate of Civilian Personnel. The appraisal system meets the

requirements of the RPR, the CSRA, the OPM, and the EEOC guidelines. The attitude toward the new Job Performance Appraisal System has generally been one of acceptance. If approached honestly, the new system provides a much more objective method of evaluation than the system it is replacing. There are some objections, chief among them is that people frequently assert that standards just can not be written for their jobs. It is not easy to write standards and some jobs are more difficult to define in objective terms than others. However, the author has found through various field tests, that standards can be written for every job. It requires a hard and honest look at the job elements and the performance level required to satisfactorily perform the job.

Other objections are that the system is expensive and time consuming. These are both valid objections; but the potential for increased productivity and fairness resulting from the system far outweighs these operational problems. It requires the education of trainers and time off the job for training of every employee and supervisor under the JPAS. It also requires some quality time and effort on the part of the participants outside of training to write work plans, review them throughout the year, and render objective ratings.

In the final analysis the success of the system lies with the participants. It can only be as good as the integrity of the people who use it.

References

- Cowan, D.K., Thompson, N.A., Guerrieri, J.A. and Vitola, B.M. Appraisal and Merit Pay Systems for Air Force Civilian General Managers. AFHRL-SR-81-36. Brooks Air Force Base, Tx.: Manpower and Personnel Division, Air Force Human Resources Laboratory.
- Guerrieri, J.A. Air Force Senior Executive Appraisal System. AFHRL-SR-81-11. Brooks Air Force Base, Tx.: Manpower and Personnel Division, Air Force Human Resources Laboratory.
- Odiorne, G.S. Management By Objectives. Belmont, California: Pitman, 1965.
- Smith, P.C., and Kendall, L.M. Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 1963, 47 (2), 149-155.
- Watson, T.W., Guerrieri, J.A., and Earles, J.A. Initial development of a job performance evaluation system (JPAS) for Air Force civilian employees. Paper presented at the Seventh Annual Psychology in the Department of Defense Symposium, USAF Academy, Department of Behavioral Sciences and Leadership, Colorado Springs, Colorado, 20-22 April 1980.

THE AIR FORCE JOB PERFORMANCE APPRAISAL SYSTEM

John A. Guerrieri
Air Command and Staff College
Maxwell AFB, AL 36112

Abstract

This paper describes the Air Force Job Performance Appraisal System (JPAS). The basic component of the JPAS is the work plan which provides a vehicle for delineating the job performance elements and performance standards of the employee's job. Each element is designated as either critical or non-critical and is assigned relative importance points. At the end of the appraisal period, the supervisor evaluates the employee's performance on each element and specifies whether he/she met, did not meet, or exceeded the requirements of the element. The supervisor then provides an overall rating based on the element evaluations. The JPAS meets the requirements of the Civil Service Reform Act of 1978 and the Office of Personnel Management.

The Air Force Job Performance Appraisal System (JPAS) that became effective 1 October 1981 is elegant in its simplicity. The system requires that the supervisor delineate the important aspects of the employee's job as well as measurable performance standards which indicate the level of achievement necessary for satisfactory performance. The evaluation of job performance against established standards minimizes subjectivity. The supervisor determines if the employee met, did not meet, or exceeded the standards, and the overall rating is based on these evaluations.

The Workplan

The workplan is the basic component of the JPAS. It provides a description of the work the employee is expected to perform during the appraisal period (the job performance elements); and a description of the required level of work (performance standards).

Job performance elements are identified as being either critical or noncritical to job accomplishment and are assigned importance points to reflect the relative importance of each of the job performance elements. At least 51% of the relative importance points must be assigned to critical job performance elements.

Performance standards are written to specify the satisfactory level of performance. They are written as practical measures so that a supervisor has no difficulty objectively determining if an employee met the requirements of the job.

THE AIR FORCE JOB PERFORMANCE APPRAISAL SYSTEM

John A. Guerrieri
Air Command and Staff College
Maxwell AFB, AL 36112

Abstract

This paper describes the Air Force Job Performance Appraisal System (JPAS). The basic component of the JPAS is the work plan which provides a vehicle for delineating the job performance elements and performance standards of the employee's job. Each element is designated as either critical or non-critical and is assigned relative importance points. At the end of the appraisal period, the supervisor evaluates the employee's performance on each element and specifies whether he/she met, did not meet, or exceeded the requirements of the element. The supervisor then provides an overall rating based on the element evaluations. The JPAS meets the requirements of the Civil Service Reform Act of 1978 and the Office of Personnel Management.

The Air Force Job Performance Appraisal System (JPAS) that became effective 1 October 1981 is elegant in its simplicity. The system requires that the supervisor delineate the important aspects of the employee's job as well as measurable performance standards which indicate the level of achievement necessary for satisfactory performance. The evaluation of job performance against established standards minimizes subjectivity. The supervisor determines if the employee met, did not meet, or exceeded the standards, and the overall rating is based on these evaluations.

The Workplan

The workplan is the basic component of the JPAS. It provides a description of the work the employee is expected to perform during the appraisal period (the job performance elements); and a description of the required level of work (performance standards).

Job performance elements are identified as being either critical or noncritical to job accomplishment and are assigned importance points to reflect the relative importance of each of the job performance elements. At least 51% of the relative importance points must be assigned to critical job performance elements.

Performance standards are written to specify the satisfactory level of performance. They are written as practical measures so that a supervisor has no difficulty objectively determining if an employee met the requirements of the job.

Job Performance Elements

Job performance elements describe the actual tasks performed on the job by the employee. These tasks are derived from the employee's position description, the supervisor's and employee's knowledge of the job, and any other document pertinent to the job. The elements should not be too detailed nor should they be so general that they can not be objectively measured. Job performance elements specify actual work tasks rather than the knowledge, skills, or abilities necessary to accomplish the job. There are two methods for entering elements on the workplan: 1) line entry method, and 2) functional category method.

The line entry format provides for the identification of a job requirement utilizing a 1- or 2-line statement. Since there are numerous jobs in the Air Force that have a limited number of relatively simple job tasks, the line entry format was proposed to meet the needs of the individuals in these jobs. Some examples of line entry job performance elements are: 1) Posts stock list changes, and 2) Reviews manuscripts for correct spelling and grammar.

Functional category elements provide a method for clustering similar tasks under a single heading. A functional category element subsumes subelements which are specified in the same manner as line entry elements. At the end of the appraisal period, each element (not the individual subelements) is evaluated. The advantage of using the functional category method rests in the concept that the evaluation of performance is based on the assessment of several subelements rather than a single line entry. Since an individual may perform tasks which are amenable to both formats, elements may be entered on the workplan using a combination of both methods. A typical functional category element might be:

1. COORDINATION: (1) Provides technical guidance to other DOD and government agencies. (2) Attends technical conferences. (3) Coordinates projects with other groups within the division.

Criticality

Each job performance element must be designated as critical or noncritical to job performance. A critical element is essential to the nature of the position, and failure to meet the requirements of the element warrants management action. At least one element on the employee's workplan must be designated as critical. For functional category elements, the designation pertains to the category and not to the individual subelements.

When specifying critical elements, the supervisor should consider such factors as: 1) the mission of the organization; 2) organizational goals; 3) sustaining the work of others (where the effort of one individual or group affects the production of another individual or group); 4) meeting

directives; 5) difficulty or complexity of the element; and 6) consequences of possible failure to perform the element satisfactorily.

Relative Importance Points

Another factor that is directly related to an individual's overall performance rating is the specification of the relative importance points for each job performance element. A total of 100 relative importance points are used to reflect the entire job. Both critical and noncritical elements are assigned relative importance points. However, at least 51 points must be assigned to critical job performance elements. The relative importance points of a functional category are assigned to the element and not to the subelements of the function.

Performance Standards

Performance standards are written to specify the level of achievement necessary for satisfactory performance of the job performance elements. A minimum of one standard must be written for each element or subelement; however, more may be written if appropriate.

Performance standards should be realistic, practical to observe, attainable, exceedable (if possible), and most importantly, they must be measurable. Realistic means that the employee can reasonably be expected to meet the requirements of the standard within the capability of the employee as he or she might be influenced by the work environment. The standards should neither be too difficult nor too easy to achieve. Practical to observe means that the rating official can actually determine if a standard has been met. All standards should be attainable for the average employee and, if possible, exceedable by the highly industrious employee in order to provide incentive to the employee. If a standard is not written in a measurable manner, then regardless of how practical or realistic the standard may be, the supervisor cannot evaluate the work performed. Standards are usually written to show a range of values to describe satisfactory performance rather than a discrete amount whenever possible. Some examples of performance standards are:

1. Stock list changes are performed at a rate of 65 to 70 changes per hour.
2. Manuscripts are reviewed and corrected 8 to 10 working days after receipt.
3. COORDINATION: (1) Guidance is initiated five working days after request. One valid complaint is permitted per quarter regarding interaction with other DOD and government personnel. (2) Results of technical conferences are documented 4 to 6 working days after the conference. (3) Each project is coordinated either verbally or in writing on a weekly basis.

Periodic Performance Reviews

The workplan was designed to be a realistic and viable document that is capable of incorporating any changes required to accurately reflect actual job requirements. Normally, revisions to the workplan are accomplished during the periodic performance reviews, however, the plan may be revised at any time. Periodic performance reviews are held at the discretion of the supervisor, but may be requested by the employee. It is anticipated that several reviews will occur during the rating period. At these performance reviews, workplan changes, training needs, and other topics relevant to the job should be discussed. These reviews provide an opportune time for the supervisor to give encouragement for satisfactory performance and to motivate the employee to even better performance. It is also an excellent time to encourage and assist employees who are falling below acceptable levels of performance to strive for improvement.

Performance Appraisal

At the end of the appraisal period, the employee's performance is evaluated against the standards established for his/her elements. Based on these evaluations, an overall performance rating is given. The performance of the employee is evaluated for each element (line entry or functional category). The supervisor specifies whether the employee met, did not meet, or exceeded the requirements of the element. An evaluation of exceeded means that the performance was better than the satisfactory level as indicated by the performance standard. An evaluation of did not meet indicates performance below the standard.

The overall performance rating is directly related to the evaluations. The overall rating is determined based on the following definitions:

SUPERIOR: Employee exceeds the performance requirements of all the job performance elements of the workplan.

EXCELLENT: Employee meets the performance requirements of all the job performance elements of the workplan and exceeds the performance requirements of the job performance elements which represent at least 50% of the relative weight in importance of the workplan.

FULLY SUCCESSFUL: Employee meets the performance requirements of all the job performance elements of the workplan.

MINIMALLY ACCEPTABLE: Employee meets the performance requirements of all critical job performance elements of the workplan, but does not meet the performance requirements of one or more noncritical job performance elements.

UNACCEPTABLE: Employee does not meet the requirements of one or more critical job performance elements of the workplan.

A rating of Fully Successful or higher enables an employee to receive a within grade increase. Excellent and Superior ratings qualify employees for performance awards such as a Sustained Superior Performance or a Quality Step Increase. These top two ratings also add years of creditable service to an employee's service date for reduction in force purposes. An Excellent adds two years and a Superior adds four years to creditable service for the next year.

A rating of Minimally Acceptable or Unacceptable requires that the supervisor take some sort of remedial action. The supervisor must provide assistance to the employee rated Minimally Acceptable in order to help him/her improve his/her performance. It is especially appropriate in this circumstance to recommend special training to improve the employee's performance. In addition to receiving assistance in improving performance, the employee rated Unacceptable must also receive career counseling. If performance continues at an unacceptable level, the employee may be reassigned, demoted or removed.

Conclusion

The Job Performance Appraisal System designed by the Air Force Human Resources Laboratory meets the requirements of the Civil Service Reform Act of 1978 (Public Law 95-454), the Office of Personnel Management, and the Air Force Directorate of Civilian Personnel. The system was designed to 1) maximize objectivity in the appraisal process, 2) enhance communication between supervisors and employees, and 3) provide a vehicle for measuring actual performance. Although the workplan development phase is difficult and time consuming, if used honestly, the new appraisal system can provide a method of enhancing the productivity of the employee and the objectivity of the supervisor's appraisal.

References

Civil Service Reform Act of 1978, Public Law 95-454.

Department of the Air Force. Performance Appraisal Program, (Air Force Regulation 40-452) Headquarters U.S. Air Force, Washington, D.C., 1 October 1980.

EFFECTS OF STRESS ON THE PERFORMANCE OF MILITARY PERSONNEL

Charles D. Spielberger, Ph.D.
Professor of Psychology and Director
Center for Research in Community Psychology
University of South Florida, Tampa

ABSTRACT

Stress is an integral part of the natural fabric of life. In common sense terms, stress refers both to the circumstances that place physical and psychological demands on an individual and to the emotions experienced in stressful situations. Any situation that involves frustration or physical danger, or in which a person's behavior is evaluated by others, can be stressful.

While stress may have positive as well as negative consequences, the negative effects generally get the most attention. Stress is widely regarded as the cause of such diverse problems as poor job performance, failure on examinations, stage fright, headaches, insomnia, skin rashes, stomach ulcers and such serious medical disorders as heart attacks and cancer. But people also become bored when the level of stress is too low, leading them to seek out stressful situations such as riding a roller coaster, skiing or watching horror films. Others take their need for excitement to greater lengths, pursuing careers as professional athletes, test pilots or astronauts.

There can be no doubt that stress plays a critical role in the performance of military personnel in both training and combat situations. Therefore, understanding stress and its effects on performance must be given a high priority by military psychologists and behavioral scientists. This is especially true for those with responsibility for conducting research on operational problems, and for selecting and training personnel for leadership roles.

In this presentation, I would like to address three main topics. I will first discuss a general conceptual framework for understanding the nature of stress and stress reactions that has evolved from laboratory and field research over the past two decades. My second goal is to describe the current status of our work on the measurement of anxiety and anger as emotional reactions to stress. Finally, I will review some recent findings on the effects of stress on the emotional reactions and the performance of Air Force and Navy recruits.

AVIATOR-AIRCRAFT INTEGRATION: ARI RESEARCH IN ARMY
AVIATION*

W. R. Bickley, W. R. Brown, J. A. Dohme, and J. H. McCracken
US Army Research Institute Field Unit at Fort Rucker

ABSTRACT

The Army Research Institute Field Unit at Fort Rucker performs research in support of three general areas within Army aviation: (1) selection, classification, and initial entry training, (2) flight simulation and performance measurement, and (3) combat unit training. Examples of work in each area to be presented are: (1) Aviator Selection: the Flight Aptitude Selection Test (FAST), (2) Flight Simulation: Training effectiveness of the prototype UH-60 Black Hawk Helicopter Flight Simulators, and (3) Combat Unit Training: Analysis of pilot tasks and training requirements for the Near-Term Scout Helicopter.

The mission of the Army Research Institute (ARI) Field Unit at Fort Rucker is to provide training research support, human resources research and technical advisory support to Army-wide aviation through direct participation with the Army aviation users. The research program spans the entire life cycle of the Army aviator from initial selection to release from active duty, with an additional focus on the reserve readiness force. There is a balanced research program of technology based and advanced development research which is product oriented and stresses solutions to operational aviation problems. Accordingly, ARI, Fort Rucker is organized in three Technical Teams: Selection, Classification, and Initial Entry Training, Flight Simulation and Performance Measurement, and Combat Unit Training.

The Aviation Selection, Classification, and Initial Entry Rotary Wing Training (IERW) Team is conducting research to develop and validate written and psychomotor tests for Army aviator applicant selection for IERW training; to assign aviators to mission specific training; to develop methods for individualized assessment and progression of trainees within self-paced multiple track IERW programs; and to develop IERW training methods and performance measurement techniques to enhance combat skills training.

The Flight Simulation Technical Team is conducting research to assist Army aviation in the development of flight simulator and training device requirements; to design and conduct evaluations of flight simulation systems such as the AH-1, UH-60, and AH-64; to perform research and development of an aviation training system to support the introduction of emerging aviation systems; and to determine maximum effectiveness of the flight simulator for training in FORSCOM units.

*The views, opinions, and/or finding contained in this report are those of the authors and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other official documentation.

The Combat Unit Effectiveness Technical Team is conducting research to develop and validate unit level training modules and assessment techniques; to investigate aviator skill loss and determine an optimum proficiency maintenance and recovery regimen; to develop self-contained training modules for night operation, mission planning, and aerial defense; and to determine the amount and nature of refresher training required to return any rated aviator to ATM performance standards after a non-flying assignment.

The scope of the research program is best presented by an overview of one example from each technical team; (1) the Army Flight Aptitude Selection Test is discussed briefly from a chronological viewpoint from its initial introduction to its current status, (2) the UH-60 Flight Simulator Training Effectiveness Evaluation is presented from transfer of training approach using the methodology from Bickley (1980) and finally, (3) the training requirements analysis for an emerging system is provided which outlines the process being used to meet the needs for aviation combat unit training early in the acquisition fielding cycle of the system.

The FAST and Aviator Selection

The Flight Aptitude Selection Test (FAST) fielded in 1966 was developed by Kaplan (1965). This FAST consisted of two batteries, one for officer applicants and one for warrant officer candidate (WOC) applicants. The FAST yielded three scores for each applicant; a Fixed Wing score, a Rotary Wing score, and a composite score. Only the composite score was used for selection. The ARI Fort Rucker Field Unit, which has responsibility for development and validation of the FAST, implemented a revised version of the FAST (RFAST) in 1980 (Eastman and McMullen, 1978). The RFAST was developed to: (1) reduce test administration time by shortening the test, (2) provide for machine scoring of answer sheets to reduce errors and (3) combine the two batteries to create a single test valid for both officer and WOC applicants.

Although a recently completed review of IERW attrition focused on minority/majority comparisons, the findings demonstrated the efficacy of aviator selection instruments in general. Altogether, 1108 officer and 2185 WOC students from FY 80 and 81 classes were studied. In general, graduates, both officer and WOCs, had higher FAST scores than eliminees. For the WOCs, the mean FAST scores of graduates and eliminees showed a decline over time. This decline was due in part to a temporary lowering of the cut score for WOCs from 300 to 270. The FAST proved to be a particularly effective screening instrument. The percentage of WOCs completing flight training drops from about 78% for FAST scores of 300 to 58% for FAST scores of 270. On the upper end of the scale, for a FAST score of 360, the percentage of successful completion rises to about 95%.

Discriminant analysis using the variables FAST, AGE, GT (General Technical Sub-test of the ASVAB) and Education level as predictors correctly classified 70-80% of the WOCs and officers as graduates/eliminees. AGE and FAST made substantial contributions as discriminants while GT and Education level did not. AGE was found to be inversely related to graduation/elimination with trainees over 30 showing a decreasing probability of completing training.

Practically speaking, the Assignment Branch of the Army Military Personnel Center, with knowledge of an applicant's age and FAST score, can make a reasonable prediction of the applicant's probability of success in flight training.

UH-60 Simulator Training Effectiveness Evaluation

Two competing prototype UH-60 Black Hawk Helicopter Flight Simulators (UH60FSs) have been installed at USAAVNC. Both are identical except for the visual system; one employs a 1000:1 scale model board and closed-circuit television system and the other utilizes computer-generated imagery. The present operational test of the UH60FS has as its main objective the determination of which visual system is most effective for training Army aviators to fly the UH-60 Black Hawk. The overall approach being taken is one of determining for each maneuver taught in each simulator the function relating amount of initial flight simulator training with amount of subsequent aircraft training.

Students at USAAVNC transitioning from some other aircraft to the UH-60 were utilized in the study. For each maneuver and each simulator, groups of students received no UH60FS training (i.e., all training in the aircraft) or one of three pre-determined levels of flight training.

Then, using amount of subsequent aircraft training required to reach proficiency as the predicted variable and amount of UH60FS training as the predictor, the data were fit to the function $f(x) = (a)(\exp(-b)(x)) + c$ (cf. Bickley, 1980). In this function, $a + c$ (i.e., $f(0)$) represents the amount of training required using the aircraft alone; c represents the amount of training that cannot be accomplished by simulation but must be completed in the aircraft; and b represents the rate at which $a + c$ decreases to c .

In general, if training costs are ignored, for any maneuver the more effective training device is the one yielding the smaller value of y (aircraft training) for some given value of x (simulator time). Likewise, a comparison of relative training effectiveness across some range of simulator training can be made between simulators by comparing integrals of the two respective functions over that range.

These results will be used both in determination of the training effectiveness of the two prototypes and also in determining their cost and training effectiveness.

NTSH Tasks and Training Requirements

A number of developments in the past decade have underscored the importance of thoroughly examining mission requirements, crewmember tasks, and training requirements concurrently with the development, testing, and fielding of new Army helicopters. Doctrinal changes which place increasing emphasis on terrain flight tactics, and a commitment to employ Army aviation assets in day/night and

marginal weather conditions have resulted in substantial increases in skill and workload requirements of helicopter aircrews. Operation at terrain-masking altitudes and/or in error, and makes it essential that flight, navigation, communication, and mission tasks be coordinated between crewmembers.

At the same time, rapidly advancing technology has given rise to a number of equipment innovations which substantially increase the versatility and effectiveness of Army helicopters. Nowhere are the equipment innovations and increased capabilities more apparent than in the NTSH operating in conjunction with the AH-64 Apache or AH-1S modernized Cobra Attack Helicopters. Highly sophisticated equipment included in these aircraft, while increasing mission capabilities, also greatly increases the skill and training requirements of the crewmembers. Effectively manning these helicopters, which will usually operate in teams, requires thorough front-end analyses for each to provide information in areas such as: crewmember task functions, task coordination requirements, crewmember aptitudes and training requirements.

This front-end analysis is presently being accomplished as follows. First, mission profiles for the NTSH operating with the AH-1, with the AH-64, or as forward artillery observer were prepared, describing mission objectives, flight modes, and targets. For each of these mission profiles, a time series analysis identifying when and for how long each piece of aircraft equipment is being operated was conducted. This analysis included for each equipment such data elements as required sensory cues, required operator cognitive processes, concurrent or sequential coordination required, and criticality of performance.

On basis of the time series analysis, equipment functions will be allocated between the NTSH crewmembers such that operator overload is minimized and success in equipment employment is maximized.

Based on these detailed task analyses, feedback for NTSH cockpit design and layout will be generated. Also, skills and knowledges required for proficient performance. All have been identified, along with recommended media for their training.

Epilog

In this short paper, a brief overview of only three of the efforts currently being engaged in by ARI, Fort Rucker has been possible. Those wanting more information on these efforts or any aspects of the ARI, Fort Rucker program are invited to contact:

Chief: C. A. Gainer (205) 255-4404 (AV 558-4404)
Initial Entry and Training: M. G. Sanders, (205) 255-3686 (AV 558-3686)
Flight Simulation and Performance Measurement: R. H. Wright, (205) 255-2873 (AV 558-2873)
Combat Unit Training: J. H. McCracken, (205) 255-2873 (AV 558-2873)

References

Bickley, W. R. Optimizing simulator-aircraft training mixes. Proceedings, 2nd Interservice-Industry Training Equipment Conference, Salt Lake City, UT, 1980.

Eastman, R. F. and McMullen, R. L. Item analysis and revision of the flight aptitude selection tests. Research Memorandum, Army Research Institute Field Unit, Fort Rucker, AL, 1978.

Kaplan, H. Prediction of success in Army aviation training. Technical Research Report 1142, US Army Personnel Research Office, OCRD, 1965.

AUTHOR INDEX

Appel, Victor H.	242, 252	Hart, Sandra G.	478
Babson, June I.R.	158	Hauser, Jan R.	478
Barkley, William M.	97	Hayes, John F.	527
Batchelor, Edward, Jr.	37	Heller, Frederick H.	106
Benel, Denise C.R.	22	Hendrix, William H.	232
Benel, Russell A.	22	Hertz, Rosanna	440
Berry, Gene A.	415	Hibler, Russell J.	209
Bessemer, David W.	83	Hoiberg, Anne	523
Bickley, W. R.	551	Hritz, Rohn J.	193, 198
Bittner, Alvah C., Jr.	69, 321	Hughes, Richard L.	237, 420
Blanding, Z. Benjamin	383	Jaeger, Bruce R.	494
Bloom, Richard W.	176, 362	Jensen, Harald E.	48
Bolton, Jose	454	Jobe, Jared B.	516
Bowman, Harry L.	57	Kane, Robert P.	462
Boyd, Stephen P.	334	Kayten, Phyllis J.	118
Boyle, James A.	494	Keeth, James B.	342
Brown, Rand L.	1	Kennedy, Robert S.	311
Brown, Paul	150	Kerr, Norman J.	57
Brown, William R.	551	Kottas, Brian L.	83
Carter, Robert C.	69	Kristiansen, Donald M.	132
Childress, Mary E.	478	Laughery, K. Ronald, Jr.	88
Clifford, Jim	170	Lawrence, Gerald H.	498, 507
Coffman, Richard M.	37	Levine, Michael	317
Coon, Kenneth C.	237	Lintern, Gavan	311
Dansby, Mickey R.	166	Longridge, Thomas	32
Derrick, William L.	473	Lord, Sharon B.	298
Ditzian, Jan L.	88	Luchsinger, Vince P.	260, 265
Dohme, J. A.	551	Majchrzak, Ann	392
Donohue, Bobby	265	Malmstrom, Frederick V.	37
Durso, Francis T.	304	Mangelsdorff, A. David	228
Dyer, Jean L.	182, 193, 198	Manley, T. Roger	203
Eddowes, Edward E.	486	Marshak, William P.	1
Emmett, Frank E., Jr.	371	Martin, Elizabeth C.	32
Evans, David W.	13	Matthews, Michael D.	415
Fairbank, Benjamin A., Jr.	317	McCloy, Thomas M.	494
Fiedler, Fred E.	283	McCracken, J. H.	551
Flores, Leona	277	McCright, Gerald D.	247
Frankle, Allan H.	352	McGill, Donald W.	193, 198
Frass, Ronald G.	507	McMenemy, J. P.	63
Furukawa, T. Paul	228	Merrin, Mary Beth	399
Fye, Samuel P.	232	Mickley, G. Andrew	498, 507
Gawron, Valerie J.	27	Mitchell, Jimmy L.	342
Gibbs, Gregory L.	498	Nadler, Larry	187
Giles, David A.	368	Nogami, Glenda Y.	403, 410
Ginnett, Robert C.	431	Ovalle, Nestor K.	232
Ginsburg, Arthur P.	13	Paap, Kenneth R.	27
Glasgow, Zita	111	Packard, Gary A., Jr.	166
Goldsmith, Timothy E.	304	Pennington, Ennis R.	77
Gregoire, Harv G.	490	Pepper, Ross L.	533
Guerrieri, John A.	540, 545	Percy, Richard L.	97
Guignard, John C.	69	Phillips, Mark D.	533

AUTHOR INDEX

Pierce, Marianne	260
Plonsky, Robert	507
Porter, David B.	158, 270
Porter, Sharon J.	270
Potter, Earl H. III	219
Ray, Jarean	150
Reamy, Christopher J.	37
Ree, Malcolm	317
Rinalducci, Edward J.	32
Roach, Bennie W.	328
Roth, J. Thomas	193, 198
Sampson, James B.	516
Schendel, Joel D.	106
Schvaneveldt, Roger W.	304
Scott, Lynn M.	42, 425
Sherrill, James L.	137
Solomon, Victor M.	468
Spielberger, Charles D.	550
Staton, Charles W.	232
Stevens, Karen E.	498, 507
Taber, Lynnvillle E.	446
Thissen, David	317
Troxler, Raymond G.	232
Vaughan, Eva D.	111
Wainer, Howard	317
Warren, James D.	252
Watson, Thomas W.	242
White, Gerald A.	498
Wiekhorst, Linda A.	342
Williams, Kent E.	118
Wilson, Chester D.	1
Kolstad, Jeffrey C.	69
Wood, Frank R.	277, 440

☆U.S. GOVERNMENT PRINTING OFFICE:1983-677-509 / 104