USAFA-TR-90-1

AD-A240 113

PROCEEDINGS
PSYCHOLOGY IN THE DEPARTMENT OF DEFENSE

TWELFTH SYMPOSIUM
18-20 APRIL 1990

91-10319

91 9 11 011

SPONSOR:

USAF ACADEMY DEPARTMENT OF BEHAVIORAL SCIENCES AND LEADERSHIP

Best Available Copy

PROCEEDINGS OF THE TWELFTH SYMPOSIUM
PSYCHOLOGY IN THE DEPARTMENT OF DEFENSE
18-20 APRIL 1990

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

SYMPOSIUM CHAIR
LIEUTENANT COLONEL DAVID PORTER

PROFESSOR AND HEAD, DEPARTMENT OF BEHAVIORAL SCIENCES AND LEADERSHIP

Colonel Richard L. Hughes

EDITOR

Capt Dana H. Lindsley

SYMPOSIUM COMMITTEES

PROGRAM COMMITTEE

Major Leray L. Leber, Chair
Major Raoul J. Buron
Major Gordon J. Curphy
Captain Gary Macomber

INVITATIONS AND REGISTRATION COMMITTEE

Major Robert N. Shearer, Chair
Major Paul Grunzke
Captain Dean H. Orrell II
Captain Andrew G. Stricker

SOCIAL, BILLETING, TRANSPORTATION AND PROTOCOL COMMITTEE

Captain Charles G. Capps, Chair
Captain Bernard W. Asiu
Captain Sharon M. Latour

FACILITIES AND VISUAL AIDS

Captain Harold S. Hillman

FINANCE

Captain Keith W. Gay

PUBLICITY

Major Karen O. Dunivin

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

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

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

Richard Durham

RICHARD DURHAM, Lt Col, USAF
Director of Research

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. Telephone (804) 487-4660.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for Public Release and Distribution Unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) USAFA TR 90-1	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Dep of Behavioral Sciences 6b. OFFICE SYMBOL (If applicable) DFBL	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State, and ZIP Code) U.S. Air Force Academy Colorado Springs CO 80840-5701	7b. ADDRESS (City, State, and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Air Force Office of Scientific Research 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
Proceedings, Psychology in the Department of Defense, Twelfth Symposium (UNCLASSIFIED)

12. PERSONAL AUTHOR(S)
LERAY L. LEBER, Major, USAF

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Proceedings	FROM ____ TO ____	April 1990	372

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB-GROUP
05	05	
05	10	

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Symposium, Training, Leadership, Counseling, Cognitive Performance, Turnover, Workload, Family Issues, Biofeedback, Retention, Women's Issues, Learning.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This printed proceedings include papers and presentaitons that deal with a wide range of reesearch in psychology with emphasis on military issues.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code)	22c. OFFICE SYMBOL
LERAY L. LEBER, Major, USAF	(719) 472-3860	DFBL

DD Form 1473, JUN 86 Previous editions are obsolete. SECURITY CLASSIFICATION OF THIS PAGE

FOREWORD

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

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

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

In publishing the Proceedings of the Twelfth 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

Proceedings & Conference Preparation:
Sandy Dunagan
Jan Huie
Laurel Mellott

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

PAGE

FEATURED SPEAKERS:

PAPER PRESENTATIONS:

Training Issues

Human Factors: General Issues

Gender Identity and Discrimination

Aircrew Issues

Human Computer Interaction

Career Intent, Retention, and Organization Commitment

Naval Recruiting: Sighting, Sounding & Course Correction

Counseling - General Issues

Performance Measurement and Feedback

Personal Development

Evaluation of Equal Opportunity Climate and Training

Personnel Selection & Job Performance

KEYNOTE ADDRESS

THE MARCH OF COCKPIT AUTOMATION: WILL HUMANS EVER REPLACE COMPUTERS?

Dr. Earl L. Wiener
University of Miami
School of Business Administration
417 Jenkins
Coral Gables, Florida 33124-8237

The rapid advances of cockpit automation the last decade outstripped the ability of the human factors profession to understand the changes in human functions required. High technology cockpits require less physical (observable) workload, but are highly demanding of cognitive functions such as planning, alternative selection, and monitoring. Furthermore, automation creates opportunity for new and more serious forms of human error, and many pilots are concerned about the possibility of complacency affecting their performance.

On the positive side, the equipment works "as advertised" with high reliability, offering highly efficient, computer-based flight. These findings from cockpit studies probably apply equally to other industries, such as nuclear power production, other modes of transportation, medicine, military systems, and manufacturing, all of which traditionally have looked to aviation for technological leadership. The challenge to the human factors profession is to aid designers, operators, and training departments in exploiting the positive side of automation, while seeking solutions to the negative side.

Mental Rotation Skills [1]

Diane L. Damos, Ph.D.
University of Southern California

Abstract

Four experiments concerned with mental rotation skills are described. The purpose of these studies was to determine the effect of extended practice on rotational skills, establish their generality, and determine if they can become automated with practice. Extended practice resulted in extremely high rotation rates. Mental rotation skills transferred between stimuli with the highest transfer between stimuli, that were most similar. The possibility of automatization could not be established.

Recently, the concept of "situational awareness" has generated much interest in both the operational and research communities. To date, most of the attempts to increase situational awareness have been concerned with changing the amount of information displayed to the pilot, the format of the information, and in some cases the nature of the displayed information.

A second approach to increasing situational awareness is through training. This approach has generated little interest to date, perhaps because of the complex, multifaceted nature of the concept. Attempting to train situational awareness as a unitary skill may, therefore, be futile. Training specific skills that contribute to situational awareness, on the other hand, may be an effective method for increasing the pilot's general level of situational awareness.

Spatial information processing skills are frequently assumed to contribute substantially to situational awareness. Although the specific skills that contribute to situational awareness have not been identified, mental rotation appears to be involved in a number of flying tasks. For example, a pilot approaching an airport from other than a southerly direction may have to rotate the airport diagram mentally to coincide with the direction of approach.

Mental rotation skills were selected for examination primarily because they appear to contribute to situational awareness. Additionally, mental rotation has been extensively studied in laboratory environments. Consequently, a large data base exists that could be used in designing experiments and interpreting results.

Four experiments are discussed briefly in this paper. These experiments were designed to provide a foundation for a training package that would increase a pilot's situational awareness by improving the pilot's mental rotation skills. The first purpose of these experiments was to determine the effect of extended practice on rotational skills. The second purpose was to determine the generality of these skills; if situational awareness is to be improved, rotational skills must transfer from a training situation to an operational setting. The third purpose was to determine if rotational skills could become automated with practice.

[1] This research was conducted under Contract No. N00014-86-K-0119. Dr. Tom Jones was the contract monitor.

General Method

Apparatus

A DEC Micro PDP 11 computer generated all stimuli, recorded and processed subject responses, and timed all trials. The stimuli were displayed on a Tektronix 4125 CRT. Subject responses were made using 4 X 4 matrix-type keypads attached to the subject's chair.

Subjects

Right-handed males between the ages of 18 and 35 were recruited for the experiment. All subjects were native English speakers and had no flight training. Subjects were paid a flat fee for completing the experiment and could receive bonuses for good performance.

Tasks

The subject's task was to distinguish as quickly as possible between a stimulus and its mirror image. Except in the training phase of Experiment 3, stimuli were presented at the upright (0 degree) position and at the 60, 120, 180, 240, and 300 degree positions. In the training phase of Experiment 3, stimuli were presented at increments of 36 degrees. The subject used the index finger of his dominant to make a "standard" response and the index finger of his nondominant hand to make a "mirror" response.

Three stimuli were used in these experiments: F, G, and a 24-point abstract shape (stimulus #29, Vanderplas and Garvin, 1959). A trial consisted of 30 representations of a stimulus and 30 presentations of its mirror image. The letter and its mirror image were presented five times at each orientation. At the end of each trial, the subject was shown his percentage of correct responses and his correct RT.

A second task, the matrix task, was used in Experiment 4. In this task 5 X 5 matrix grids were presented sequentially to the subject. Each matrix had five, randomly-selected illuminated cells. The subject's task was to determine as quickly as possible if the current matrix was identical to the preceding matrix. If the current matrix was identical to the preceding matrix, it was not presented in the same orientation; the second of two identical matrices was always rotated 90 degrees relative to first matrix. If the current matrix was a rotated version of the immediately preceding matrix, the subject responded *same*. Otherwise, he responded *different*. Approximately 50% of the correct responses on any given trial were *same* and 50% *different*.

Procedure

Subjects in the experimental groups in Experiments 1, 2, and 3 completed one testing session per day on each of ten working days. These subjects performed the mental rotation task with one stimulus during the first week and with the second stimulus during the second week of testing. Subjects in the control groups of these three studies received only five experimental sessions. Each subject in the first three experiments made 10800 responses each week of the experiment.

Subjects in Experiment 4 received differing amounts of practice. The procedure for this experiment is described in the appropriate section below.

Experiment 1

Introduction

The purpose of this experiment was to examine the effect of extended practice on the mental rotation task and to begin examining the generality of the required skills. Two letters, F and G, were chosen as stimuli. Because both of these letters were familiar to the subjects and both were the same type of stimuli (letters), this experiment was expected to produce the greatest number of statistical tests demonstrating positive transfer.

Twenty subjects were divided into two groups. One group trained with the letter F and transferred to the letter G. The second group trained with G and transferred to F. Although there was no true control group in this experiment, the transfer performance of one group could be compared to the training performance of the other group.

Results

Preliminary analyses. To understand the results, the analysis procedure must be understood. Correct responses in the mental rotation paradigm are related to the position of the stimulus in the following way:

$$\mathbf{RT = alpha + beta\ (D)} \qquad (1)$$

where:
RT is the correct reaction time
alpha is the intercept of the equation
beta is the slope of the equation
D is the absolute degrees of rotation of the stimulus from vertical

Thus, mental rotation data can be described by two parameters, alpha and beta. These two parameters are associated with specific cognitive processes (see Berg, Hertzog, and Hunt, 1982 for a brief discussion). The slope, beta, reflects the speed of the mental rotation of the stimulus. The intercept reflects the encoding of the stimulus, comparison of the encoded stimulus with an image held in memory, and response selection and execution. Thus, because data from the mental rotation paradigm can be represented by Equation 1, the effects of extended practice and timesharing on rotational skills can be determined separately from their effects on the encoding, comparison, and response selection stages. Similarly, transfer of rotational skills can be examined separately from the transfer of the other cognitive processes.

For the first three experiments, each testing session was divided into two parts. The raw data within each block were pooled to create an average correct reaction time (RT) for the standard and mirror versions of the stimulus at each absolute degrees of rotation for each subject. Equation 1 was then fit to the data from each part. Several analyses, which are described below, were conducted on these data.

For a variety of reasons, estimates of transfer could not be obtained using traditional techniques. Instead, the method of calculating transfer of training using curve fitting described by Spears (1985) was adopted. The data from the first three experiments were fit by an exponential equation of the form:

$$\mathbf{dv = c\ exp\ (gx) + h} \qquad (2)$$

Where:

c, g, and h are the parameters to be fit
x is the block number
dv is the dependent variable (alpha or beta).

Each of these three parameters represents different aspects of human performance. The rate of improvement in performance is represented by the second parameter, g. Asymptotic performance, the performance level after an infinite number of trials, is estimated by the third parameter, h. The sum of the first and third parameters (c + h) represents the point at which the function crosses the y-axis. In human performance terms, the quantity (c + h) represents the level of performance before any practice occurs on the task (Spears, 1985) and is known as the beginning level. Thus, this technique provides a more detailed analysis of transfer of training than the traditional formulas, which result in only one global estimate of transfer.

Estimates of transfer. The results of all of the transfer analyses are too numerous to recite. Briefly, analyses conducted on beta, the rate of rotation, showed positive transfer for both groups on c, for one group on g, and for the other group on h. Thus, prior practice affected the beginning level of the rate of rotation, the rate of learning, and its asymptotic level. To provide the reader with some idea of the levels of performance achieved, the estimated asymptotic rate of rotation for one group was .37 ms/degree, which is equivalent to 2700 degrees/s.

Analyses conducted on alpha again showed transfer benefits for the c, g, and h parameters. Thus, prior practice benefited the initial duration, rate of improvement, and asymptotic level of the durations of the encoding, comparison, and response selection stages of processing. Again, differences in the asymptotic level were unexpected.

Effect of extended practice. Because the average rate of rotation was so great, the subjects may effectively have ceased to rotate the image mentally before making a response. To test the idea, two three-way analyses of variance (ANOVAs) (Group X Stimulus Type X Rotation) were conducted on the correct RTs of the last block of trials of each week. Neither ANOVA showed a significant effect of rotation.

Discussion

To interpret the results, it is easiest to begin with the transfer data. The significant between-group differences on c for both alpha and beta were not surprising; initial performance on a transfer task typically benefits from prior practice. The transfer effect on g for alpha and beta also was expected. The effect on h for both alpha and beta, however, was surprising; 5 hours of practice was not considered sufficient to produce differences in the estimates of asymptotic performance between the "experimental" and the "control" groups. These results cannot be explained at present.

The positive transfer for alpha may have fewer practical implications than apparent. As noted earlier, alpha represents three stages of processing. Theoretically, positive transfer could result from changes in the duration of any combination of the three. In this experiment, the most logical source of this change is the response selection stage. Before the subjects began the transfer phase of the experiment, they made almost 11000 responses with the keypad. Consequently, the duration of the response selection stage should have been minimized by the end of the training phase. Because the same keypad was used in both phases of the experiment, the duration of the response selection stage may have been minimal at the beginning of the transfer phase, resulting in the positive transfer for the estimated initial level of performance. The low initial value of c may also have caused the apparent difference in g, the rate of improvement parameter.

The data indicate that with extended practice, the correct RT was no longer related to the angle of rotation. Mathematically, this result implies an infinitely high rate of rotation. From a more practical perspective, this finding might indicate that performance became automated with practice. No operational definition of automaticity exists for a mental rotation task. Consequently, these results are suggestive of automaticity but not definitive.

Experiment 2

Introduction

Although the results of the first experiment were encouraging, the use of a strictly within-subject design limits the conclusions that can be drawn from the data; asymmetric transfer can occur in such designs, influencing the data in unpredictable ways. The second experiment attempts to replicate the results of Experiment 1 by using a traditional transfer of training design. Such designs are not susceptible to asymmetric transfer. The second experiment provides more insight into the generality of mental rotation skills by examining transfer from letters to an abstract shape.

Thirty subjects were divided into three equal groups. One group was an experimental group that trained with the letter G. Another was an experimental group that trained with the letter F. The third group, a control group, participated only in the transfer phase of the experiment, which used the abstract shape.

Results

Estimates of transfer. Analyses performed on response curves for beta showed no significant between-group differences. Analyses performed on alpha showed between-group differences favoring the experimental groups on c and h, the initial level and the estimated asymptote.

Effect of extended practice. The same analysis used in Experiment 1 was conducted on the transfer data from this experiment. These data showed an effect of rotation, indicating that even with extended practice, correct RT was related to the angle of rotation.

Discussion

The results of this experiment were considerably less encouraging than those of Experiment 1; prior practice with a letter had no effect on the rotation rate for an abstract shape although the initial level and the asymptotic level of alpha still benefitted. Again, the positive transfer for alpha probably reflects familiarity with the keypad and is of little practical significance. Additionally, extended practice did not alter the relation between correct RT and the angle of rotation as expressed in Experiment 1. Thus, the tentative evidence for automaticity found in Experiment 1 was not replicated in Experiment 2.

Experiment 3

Introduction

Although less transfer was found in Experiment 2 than in Experiment 1, in both experiments the training and the transfer stimuli were presented in increments of 60 degrees. Thus, the observed transfer may have been caused by familiarity with the position of the stimuli rather than by transfer of rotational skills. Experiment 3 used different degrees of rotation in the training and the transfer phases to eliminate any spurious effects caused by familiarity with the position of the stimuli.

Twenty subjects were divided into two groups of ten subjects. One group was an experimental group that trained with the letter G. The second group, a control group, participated only in the transfer phase of the experiment, which used the abstract shape.

Results

Estimates of transfer. The experimental and control groups differed significantly on c and h for beta. Both analyses demonstrated better performance for the experimental group than for the control. The groups did not differ on alpha.

Effect of extended practice. As in Experiment 2, an analysis conducted on the transfer data showed that extended practice did not affect the relation between correct RT and the angle of rotation.

Discussion

Experiment 3 provided evidence for the transfer of rotational skills that cannot be attributed to the specific degrees of rotation used for the training and transfer stimuli. More importantly, transfer was found for beta, which reflects the rate of rotation, rather than for alpha, which reflects the encoding, comparison, and response selection stages. Although the rate of improvement in the speed of rotation (g) was not affected by prior practice, both the initial level of performance and the asymptotic level benefitted from prior practice.

Experiment 4

Introduction

This experiment examined dual-task mental rotation performance as a function of practice and had two primary purposes. The first purpose was is to determine if performance on the mental rotation task can become automated with practice. The results of Experiment 1 provided some indication that performance may have been automated. Experiment 4 addresses the question of automaticity directly by having subjects perform a second task concurrently with a well practiced mental rotation task. If performance on the mental rotation task is unaffected by the presence of another task, then the mental rotation task can become automated.

The second purpose of the experiment was to determine if dual-task performance is affected by the amount of prior single-task practice. Because many types of timesharing skills may be involved in dual-task performance, single-task practice beyond some minimal amount may have little effect on dual-task performance. Three levels of single-task performance were included in Experiment 4 to determine if increasing the amount of single-task practice beyond some minimal amount affected dual-task performance. The three levels required 10800, 6480, and 900 responses before the dual-task phase of the experiment. One group of ten subjects experience each of the levels of single-task practice.

Results

The results of this experiment are very complex and will be discussed only briefly. To understand the results of the dual-task analyses, the single-task performance levels of the three groups on the mental rotation task must be described. The group with the most extensive practice had the fastest correct RTs but the highest error rate. The group with the least practice had the longest average correct RT and an intermediate error rate. The three groups received the same amount of practice on the matrix task and did not differ on either correct RT or error rate. The dual-task analyses revealed no clear advantage for the group having the most extensive single-task practice and did not reflect the single-task rank ordering of the groups.

An ANOVA was conducted on rotation task data for the group that received the most extensive single-task practice. The ten trials preceding dual-task practice were analyzed. The analysis revealed that correct RT was related to the angle of rotation.

Discussion

The results indicate clearly that increasing the amount of single-task practice beyond some minimum does not improve dual-task performance; the between-group differences did not uniformly favor the group that received the most single-task practice. The data were, however, considerably less clear concerning automaticity; the single-task data indicated that correct RT was affected by the angle of rotation. Thus, these data did not reach one plausible criterion for automaticity. The most common dual-task criterion--lack of a decrement--also was not met. Thus, performance on the mental rotation task does not appear to become automatic with practice.

General Discussion

As noted in the Introduction, these experiments were designed to provide a foundation for a training package to increase situational awareness. On the whole, the results of the first three experiments were encouraging. Mental rotation skills improve dramatically during the first 5 hours of practice. These skills also appear to transfer between stimuli and between specific presentation positions. Some of the transfer may be caused by familiarity with the response devices, but some of it appears to be the result of improved skills in two-dimensional rotation.

Interpreting the results of the fourth experiment proved problematic in part because no generally accepted operational definition of automaticity for a mental rotation task exists. Extended single-task practice had no obviously beneficial effect on dual-task performance, and providing more than a minimal amount of single-task practice may not be cost effective.

More applied research must be done on mental rotation skills. Clearly, additional work on three-dimensional rotation is needed. At some point, rotation should be studied in conjunction with movement of the object. The results of these studies indicate that training mental rotation skills may be a cost-effective method for improving situational awareness.

References

Berg, C., Hertzog, C., and Hunt, E. (1982). Age differences in speed of mental rotation. Developmental Psychology, 18, 95-107.

Spears, W. (1985). Measurement of learning and transfer through curve fitting. Human Factors, 27, 251-266.

Vanderplas, J., and Garvin, E. (1959). The association value of random shapes. Journal of Experimental Psychology, 57, 147-154.

INVITED SPEAKERS

ALCOHOL AND SOCIETY

Gary G. Forrest, Ed.D., Ph.D.
3208 N. Academy Blvd, Suite 160
Colorado Springs, CO 80917

The focus of this presentation will include: (1) a brief review of the historic uses/abuses of alcohol in different societies, (2) epidemiology and patterns of alcohol use/abuse in modern American society, and (3) an examination of the various parameters of responsible drinking/alcohol use in modern American society.

ORGANIZATIONAL FACTORS IMPACTING TEAM PERFORMANCE

Robert C. Ginnett, Ph.D.
Department of Behavioral Sciences
and Leadership
USAF Academy, CO 80840-5701

In 1983, Richard Hackman (now at Harvard University) and I embarked upon a long term NASA funded research program. The purpose was to understand and improve team performance in aeronautical and space environments. That purpose has remained unchanged but the funding and focus have altered as we have learned more about the phenomenon and also as we grappled with the areas we did not know enough about. For the current phase of the research, total funding comes from the "space" side of NASA and is tied directly to our realization that long duration space flight will require a better understanding of groups at work. Afterall, three years together in a confined capsule on a Mars mission will necessitate some group process skills if the team is to avoid self-destruction.

Our focus has evolved more as a function of our learnings in earlier phases. In the first phase, I examined in detail the formation process of commercial airline flight crews and, at a coarser level, their on-line crew performance. A most important finding was the power of the leader in forming an effective crew from the first moments of the crew's life. This study also resulted in a few refinements of Hackman's normative model of team performance, particularly in the area of authority dynamics associated with effective vs. less effective captains.

We also realized that teams forming and operating in organizational settings were very different from teams of students or groups of psychiatric patients in therapy. In fact, the organization and its environment played a critical role in establishing the conditions for leadership and team effectiveness. We have labeled these conditions "organizational shells." But to understand which organizational factors are most appropriate for enhancing team effectiveness would require a large scale research effort--one that looked at teams performing the same type of work but across organizations (as opposed to within the same organization). It also would require teams of researchers.

After a brief review of the background, the presentation will describe the methodology of Phase Two of the research, a general description of the types of organizations examined, and a brief discussion of the data collection procedures and the teams that collected the data, including the range of problems encountered. Finally, a presentation of the preliminary results of the organizational factors based upon the perceptions of observed crews will be presented. Opportunity for discussion and questions will be provided.

TOWARD AUTOMATING COGNITIVE-ORIENTED TASKS

Richard Koubek, Ph.D.
College of Engineering & Computer Science
Dayton, OH 45435

During the early part of the 20th century, the industrial revolution brought about radical changes in the workplace by implementing automation technology for physical task components of a worker's job. Society is presently in the throes of another revolution with the advent of the information age, where cognitive task components are now being automated with such technologies as expert systems and neural nets. While many parallels between the two periods exist, the information age is unique in that task components being modified are no longer visible and must be inferred.

Due to the complexity of interesting cognitive tasks, and current technological limitations, "brute force" approaches have only been marginally adequate. Rather, respected members of the field have suggested that in order to obtain performance at the human expert level, systems must necessarily incorporate not only the content, but also the manner, in which human experts perform the cognitive task. Therefore, understanding highly skilled cognitive performance is a critical element in the effort to automate cognitive tasks. The present literature suggests at least three components contribute to cognitive task performance: automation, problem representation and cognitive abilities (individual differences). Koubek and Salvendy (1988) performed a series of studies examining these approaches and have found knowledge representation to be the primary determinant in high level cognitive performance.

In this presentation, the industrial and information revolutions will be contrasted to identify parallels and potential points for leveraging automation into cognitive tasks. Second, it will provide an overview of approaches to skilled cognitive task performance and indicate how they can serve as a basis for expert systems development. Finally, research will be presented which attempts to determine how the approaches can be combined to serve as a basis for developing automation technology for cognitive-oriented tasks.

STAYING OFF THE GAME TRIANGLE AT WORK AND HOME

Thomas P. O'Hearn, Ph.D.
Front Range Institute
425 West Rockrimmon Blvd.
Colorado Springs, Colorado 80919

Open, honest sharing of thoughts and feelings can be one way of defining "intimacy". In transactional analysis theory, there are six ways of spending time. They are: withdrawal; rituals; pastimes; activities; games; and intimacy. In order to achieve intimacy the essential ingredients are two or more individuals who are seeing themselves and the others with whom they are interacting as being "OK". When a "not OK" position enters into the transaction then intimacy is not available.

Getting on the "game" or "drama" triangle can make for nonproductive work as well as personal relationships. Learning ways to stay out of "gaming" transactions can increase productivity by creating an atmosphere where individuals are free to share their ideas and feelings. A theoretical model will be presented which will look at how philosophical life positions can enhance or interfere with transactions.

MANAGING BY METAPHOR: A NOVEL APPROACH TO LEADERSHIP

Barrie Richardson, D.B.A.
Dean of the School of Business, Centenary College
Shreveport, Louisiana 71104

How can mediocre organizations be transformed into extra-ordinary organizations? What does it mean to be extraordinary? Are there principles, concepts, or guidelines that can be used to get extraordinary performance from ordinary people? What does it take to do common things uncommonly well?

Leaders of outstanding organizations in any field of human endeavor seem to have certain things in common. For one thing, they do not use complicated, abstruse, or theoretical techniques. Rather, they use simple, practical concepts. These deceptively simple ideas, like the ideas contained in the Sermon on the Mount or the Bhagavad-Gita, are easy to comprehend, but not always easy to put into action.

An idea can be simple and yet not be simplistic. The Jaguar has simple yet elegant lines; Einstein's theory of general relativity is summarized in a simple statement. Adam Smith's metaphor of people following their own self-interest as if they were being led by an invisible hand to serve the common good is both a simple and brilliant insight that captures the essence of how the free market works.

Simple yet elegant management ideas are presented as metaphors. By metaphors, I mean a variety of anecdotes, parables, analogies, similies; in short, the kind of text that conjures mental pictures that can be easily stored, retrieved, and communicated. This is, of course, an ancient approach to passing on information, but one worth harking back to. For in trying to become modern, objective managers, we often squeeze out the juice, the poetry, the rhetoric, the feelings, and the emotions that are, to a great degree, what human activity is all about.

This presentation will present five examples of powerful metaphors which can help galvanize a team and energize individuals to perform at an exceptional level. The metaphors are:

1) The 10% Principle,
2) Metaphor of the Redwoods
3) "The Chinese Eye"
4) Easter Island Mystery
5) The Beaver Paradox

This is an unusual approach to management, but there is evidence that both college students and managers from all industries find the approach memorable and useful.

HISTORICAL STUDIES IN LEADERSHIP

Larry Silverman
University of Colorado
Boulder, Colorado

If we accept the principle that leadership can't be taught in the classroom, then we must look elsewhere for ways to guide the development of leaders. Historical accounts contain many individuals whose character, development and acts teach us important lessons about leadership. By starting with Plutarch's biography of Alexander the Great and progressing through leaders such as Caesar, Charlemagne and Joan of Arc, certain broad patterns become salient. Studying figures from recent history such as Ataturk, Churchill and de Gaulle reveals how these patterns have continued to develop. My approach aims to sensitize students to the qualities and behaviors of acknowledged leaders from the past so they might better be able to recognize these qualities in themselves and their contemporaries.

LEADERSHIP EDUCATION FOR UNDERGRADUATES

RICHARD L. HUGHES, Ph.D.
Department Head
Behavioral Sciences and Leadership
USAF Academy

Many successful civilian and military leaders believe leadership is not something one can learn in a classroom. It develops through actual experience, so the argument goes, not from bookwork. This view draws too black and white a distinction between experience and study, but those of us in academia are guilty of a different error - sometimes claiming more transferability of intellectual lessons to "real life" than may be justified. In any case, both groups have been asking the wrong question. Rather than "How much impact do academic courses have on a student's subsequent leadership?" we need to ask "How do academic courses impact a student's subsequent leadership development?" We need to think more about how coursework can contribute to a life-long continuing process of discerning critical lessons about leadership from experience. That orientation affirms the value of actual experience in leadership development, but it also affirms the value of conceptual sophistication in being able to learn from experience. That orientation also suggests alternative pedagogical approaches to the leadership curriculum that would be geared toward potentiating the student's readiness to learn from experience.

LEADERS UNDER STRESS: COGNITIVE MANAGEMENT IN REVOLUTION AND WAR

Peter Suedfeld, Ph.D., F.R.S.C.
The University of British Columbia
#235-2075 Wesbrook Mall
Vancouver, B.C. Canada V6T 1Z3

Critiques of high-level decision making in international and national crises have frequently focused on the errors that leaders commit in processing information. Such errors, usually in the direction of oversimplifying the cognitive field (e.g., ignoring relevant information, discarding possibly fruitful options too soon, not considering how a particular decision would be perceived by the other side), lead to hasty and inadequate decisions, and outcomes ranging from the less than optimal to the disastrous. Such concepts as "satisficing," "groupthink," "symptoms of deficient decision-making," and the "cognitive miser" model all accept this paradigm to a greater or lesser extent.

An alternative, the "cognitive manager" model, holds that information processing tends to be as simple or as complex as the circumstances appear to warrant. Good decision makers are those who accurately assess the appropriate level of complexity, neither skimping attention and cognitive effort when those are crucial nor expending unnecessary time and energy on decisions that do not require it. Such individuals, in any walk of life, are good cognitive managers (as opposed to cognitive misers, who consistently use minimum effort regardless of circumstances).

A reliable quantitative measure of integrative complexity (the decision maker's perception of several dimensions and perspectives on problems and formulation of solutions that incorporate a number of these), has been applied to historical documents to test the cognitive manager model among political, military and revolutionary leaders of many countries through a span of several centuries. Both personal and occupational pressures tend to lead to reduced integrative complexity; in fact, such reductions consistently precede simplifying solutions to major problems (e.g., the decision to go to war). Successful decision makers are those who can recognize when there is a need for moderately high levels of complexity, and to adjust or maintain their complexity levels accordingly, even under severe stress. Rather than exhorting or training leaders to strive for maximal complexity at all times, a good strategy for leadership education should make them aware of the cost-benefit tradeoff between simple and complex information processing in particular situations.

Panel

Equal Opportunity in the Department of Defense: Looking Toward a New Century

Lt Col Mickey Dansby, DEOMI Director of Research
Jerry Scarpate, Chief of DEOMI's Research Division
Dr. Dan Landis, Director of the Center for Applied Research and Evaluation, University of Mississippi
Dr. James Thomas, consultant (formerly of Army Research Institute)

The Defense Equal Opportunity Management Institute (DEOMI) is responsible for training equal opportunity (EO) advisors from all Services and for conducting research on EO issues. This panel session will explore trends in the military EO arena. DEOMI representatives will present the latest data on EO within the DoD, highlighting trends in representation and utilization of women and minorities. Issues of concern as we enter a new century (e.g., ethnic imbalances between officer and enlisted corps, ethnic/gender composition of the future recruiting pool, sexual harassment, and combat exclusion for women) will be discussed. Panelists will discuss trends and policy changes that may influence EO in the future. Implications of the changing demographic composition of the Armed Forces will be considered. Current research and policy plans will be discussed as a springboard for audience interaction with the panel. A number of DEOMI publications will be available for the audience.

Panel

"Lean and mean, but not too short..."

H. Wallace Sinaiko, Panel Chair
Dr. James Vogel, U.S. Army Institute of Environmental Medicine
Dr. Joe McDaniel, AAMRL/HE, Human Engineering Division
Dr. Arthur L. Korotkin, University Research Corporation
Bethesda, Maryland
Dr. Harold P. Van Cott, Committee on Human Factors
National Research Council, Washington, D.C.
Dr. W. Steve Sellman, OASD, The Pentagon, Washington, D.C.
Dr. Dave Robertson, Navy Personnel Research and
Development Center, San Diego, California

The panel will address an important and neglected area of manpower research and policy, i.e., physical standards for military occupations. Although the U.S. services have long had medical fitness standards for prospective entrants, relatively little is known about how such standards are related to the actual tasks required of military personnel, or whether those standards are valid. The domain is of increasing importance because, in this prolonged period of demographic decline in the prime recruiting market, the Department of Defense may be forced to make new choices vis-a-vis entry criteria. That is, in order to maintain a quality force we may have to seek recruits who would heretofore, have been ineligible for service. Physical or medical fitness policies may have to be changed in order to provide needed recruits. Further, it may become necessary to redesign equipment to accommodate physical characteristics of service entrants. Change should be based on established fact rather than historical precedent and prejudice.

Panelists will deal with such issues as: current physical selection standards, methods for developing standards and guidelines, the relationship of those standards to the performance of military jobs, human factors design for "non-standard" operators, gender and ethnic factors, corrective treatment, and political and societal factors. Participants are knowledgeable concerning military personnel specialties and related occupations in the private sector. We expect the outcomes of the meeting to be: a) an agenda for research, and b) actionable recommendations that can be implemented without further research and development.

Panel

Clinical Consultation within Military Operational Units: Trends and Prospects

Edwin L. Gerwell, Behavioral Analysis Service,
Department of Psychology,
Wilford Hall USAF Medical Center
Edna Fiedler, Behavioral Analysis Service,
Department of Psychology,
Wilford Hall USAF Medical Center
A. David Mangelsdorff, Health Services Command,
Fort Sam Houston, TX
Walter Sipes, School of Aerospace Medicine,
Brooks AFB, TX

Traditionally the arena of organizational consultation is viewed as the applications area for industrial, organizational psychology. However, elements such as common base of training, limited professional resources and preventive applications of psychology, have tended to move clinical psychology into the organizational consultation role. Within the military, clinical psychologists tend to be larger in numbers, in proximity to operational organizations and dispersed due to assignment to base medical units. Therefore, they are potentially more responsive to local demands for psychological services.

The various services have at times systemically addressed demands of operational organizations by creating specialized mental health units. More commonly, individual psychologists have responded to local demands and their own interests. Examples of the former are the Army's division psychologists. Pilot or military police stress management programs offered to local squadrons represent the latter. The scope and effectiveness of these various involvements have been inadequately documented and explored. There is reason to believe this process may portend a greater integration of the various specialties of psychology.

The purpose of this panel is to explore various elements of this process of clinical involvement in military organizational consultation and its implications.

CATASTROPHIC DEATH IN MILITARY SETTINGS: WHAT ARE THE NEEDS OF THE SURVIVORS?

LTC Robert K. Gifford, Ph.D., U.S. Army
LTC James E. McCarroll, Ph D., U.S. Army
CPT Paul T. Bartone, Ph.D. U.S. Army

Walter Reed Army Institute of Research

Mary P. Tyler, Ph.D.
MAJ Robert J. Przybelski, M.D., U.S. Army

U.S. Army Medical Research Unit - Europe

When unexpected deaths occur in a military unit, there is great stress on the survivors, who are expected to continue to perform their military duties while resolving their grief and coming to grips with memories of events that evoked shock and often terror. Researchers from the Walter Reed Army Institute of Research and the U.S. Army Medical Research Unit - Europe have studied the recovery of units and individuals after fatal training accidents, aircraft crashes, and combat. Findings regarding the special needs of different groups of survivors - including friends of the deceased, unit leaders, medical personnel, body handlers, and casualty assistance personnel - are discussed.

Stress on Military Leaders after Sudden Death in the Unit

LTC Robert K. Gifford, Ph.D.

Leaders of military units that have just suffered fatalities must attend to the morale of surviving members of the unit while restoring the unit to full mission capability. Further, they typically have to do this in the midst of investigations and while under public scrutiny. Frequently this creates a situation in which the leader ensures that everyone in the unit gets excellent social and psychological support, but there is no one to support the leader. The lack of support for leaders stems both from a self-imposed ethic of setting a strong example and from the fact that in many settings the leaders have no local peers to provide social support. Leaders develop a variety of mechanisms to cope with this, including learning to draw strength from their soldiers, talking with chaplains, and confiding in their spouses. Research suggests that their are several ways in which the psychologist, or other consultant, can assist the leader without being intrusive or interfering with the unit's recovery process.

Medical Pers...nel Stress After Fatal Military Training Accidents

MAJ Robert J. Przybelski, M.D.

Medical personnel responding to fatal military training accidents experience extreme personal and professional stress as they attempt to revive crushed or dismembered buddies while superiors and friends look on. Our studies of five such accidents have found that support, both personal and professional, is critical for the well being and continued effective function of the medical personnel after these experiences.

Medical corpsmen received personal support from unit members and peers also grieving the loss. The professional support that helped the corpsmen cope with the death of their patients came by way of medical debriefings provided by their supervising physician assistants.

In contrast, the physician assistants had no peers locally to assist them through the grieving process. Regarding their professional needs, no patient follow-ups, feedback on care provided, or autopsy reports were afforded them by the accepting medical facility physicians. This lack of support prompted an overly critical self-evaluation and unfounded concerns regarding the appropriateness of the care they provided. The resulting loss of self-assurance adversely affected the physician assistants' quality of life as well as their ability to provide health care.

Recommendations to supervisors of these health care providers include the following: provide a medical debriefing for each level of provider as soon as possible after the fatal training accident experience; keep the forward medical personnel informed of the status of the patients they cared for, including the findings of any autopsies performed.

Fatal Training Accidents: Bereaved Friends

Mary P. Tyler, Ph.D.

Close friends of deceased soldiers experience a significant and painful loss. They also receive strong support from peers and co-workers, and can play a key role in the unit's recovery. Data from a study of four fatal training accidents in Army combat units suggest that the primary need of close friends is to be left with their unit, so that a naturally occurring healing process can take place.

Within the framework of work groups and friendship groups, the needs of bereaved soldiers are usually recognized very quickly by their peers and leaders. Initially, those most affected are held and comforted. Later, they receive visits from less bereaved soldiers, and talk together in informal groups to work through their grief and trauma.

Informal group rituals such as making scrapbooks for the deceased soldier's family also contribute to recovery. The memorial ceremony and other formal rituals are helpful in providing a sense of closure, an opportunity to express strong feelings in an appropriate environment, and an affirmation of group unity and shared values.

Besides receiving support from other unit members, the most bereaved soldiers have a unique opportunity to facilitate unit recovery. The most bereaved are perceived as having a special license to forgive those troubled by guilt, e.g. unsuccessful rescuers, and to urge the unit to return to duty in honor of the deceased. Unit members find formal rituals more meaningful and comforting if the most bereaved play a significant role.

Addressing the Needs of Body Handlers Following Mass Death

James E. McCarroll, Ph.D.
Department of Military Psychiatry
Walter Reed Army Institute of Research
Washington, D. C. 20307-5100

Robert J. Ursano, M. D.
Professor and Associate Chairman
Department of Psychiatry
Uniformed Services University of the Health Sciences
Bethesda, MD 20814-4799

Military personnel can be called upon to provide assistance when mass death occurs. The circumstances of such calls for assistance are almost always unexpected and people are called upon to perform tasks for which they have had no training or preparation. The presentation will be oriented to actions and decisions that need to be considered by command in the event that a military unit or installation is involved in the processing of bodies. This paper will describe some activities of military personnel following recent disasters and the reactions of some people to their experiences. We will describe the major stressors that are likely to be encountered, the range of reactions of people during and after such an event, and actions that might be considered by command in the form of a brief disaster human resources management plan.

MACROERGONOMICS: A SYSTEMS APPROACH FOR IMPROVING THE EFFECTIVENESS OF MILITARY SYSTEMS

Hal W. Hendrick, Ph.D.
Ogden Brown, Jr., Ph.D.
University of Denver

Andrew Imada, Ph.D.
Najmedin Meshkati, Ph.D.
University of Southern California

Summary

The concept and basic dimensions of macroergonomics are reviewed. Macroergonomics methods such as organizational and work systems analysis and participatory ergonomics are reviewed. U.S. and international examples are presented from the panelists' research and experience. Macroergonomics issues and potential applications in the miliary environment are discussed.

PANEL PROPOSAL:

APPROACHES TO STUDYING CREW PRODUCTIVITY AND GROUP PROCESSES USING HIGH FIDELITY FLIGHT SIMULATORS

In recent years, the use of flight simulators in aviation training has expanded (Foushee, Chidester & Helmreich, 1988). Flight simulators, however, also have applications for research since they provide a way of studying detailed behavioral sequences and communication patterns which would not be possible in actual line operations. The development and use of full mission simulator scenarios could aid researchers in examining, under controlled conditions, such things as crew productivity, group processes, crew coordination and leadership (Foushee, et al. 1988). The proposed panel will attempt to highlight the various ways high fidelity simulators can be used to study aspects of aircrew productivity and coordination.

Each panelist has conducted research using data from a high fidelity B-727 aircraft simulator in a NASA-Ames simulation facility. The panel members were selected because they illustrate different methods of using flight simulators to examine crew productivity issues. For example, Tom Chidester takes a fairly traditional experimental design approach to examine the relationships between input variables (i.e., personality) and outcomes (i.e., measures of team performance); Steve Predmore, Barbara Kanki and Sharon Conley, in contrast, use more exploratory methods to examine the group processes that mediate such relationships. Regardless of the specific focus or approach, the method of using simulator data to examine aspects of crew processes and performance is common across presenters. The panel session will be designed to encourage comments and discussion from the audience. **Robert Ginnett** will serve as chair of the session. The panelists are as follows:

1. **Thomas R. Chidester (NASA-Ames, Moffett Field).** Tom Chidester will present a simulator study examining the impact of crew selection along dimensions of personality on crew effectiveness. The study design includes the use of a "cluster analysis" method to cluster personality traits (Instrumentality, Expressivity and Achievement Striving) into general constructs. This simulator study provided a database for the more exploratory studies of group processes conducted by Predmore, Kanki and Conley.

2. **Steve Predmore (University of Texas, Austin).** Steve Predmore will discuss the design and findings of his study examining how crews attend to abnormal operations and interruptions during simulated flight. He uses an extreme-case method, contrasting the most and least effective crews.

3. **Barbara G. Kanki (NASA-Ames, Moffett Field).** Barbara Kanki will discuss the use of a simulator design to examine overall communication patterns that differentiate more and less effective crews. The study focuses specifically on interactive behavioral sequences and the way in which communication patterns mediate the effect of personality on group performance.

4. **Sharon C. Conley, (University of Arizona), Don Bryant (NASA-Ames, Moffett Field), and Yvonne Cano (University of Arizona).** Sharon Conley, Don Bryant and Yvonne Cano will report the design and findings of their study examining crew dynamics of more and less effective crews. The research contrasts four three person crews flying three legs of flight in an aircraft simulator. The study specifically examines informal group processes that facilitate task coordination.

5. **Kathy Mosier O'Neill (University of California at Berkley).** Kathy Mosier O'Neill will discuss team and expert decision making strategies, use of information, and the influence of personality on the decision making process.

Panel Discussion of Aircrew Performance Measurement Issues

Wayne L. Waag
Herbert H. Bell
Air Force Human Resources Laboratory

Richard W. Obermayer
Donald Vreuls, Panel Discussant
VRC Corporation

Anthony P. Ciavarelli
Navy Postgraduate School

Marty R. Rockway
Gary S. Thomas, Panel Chair
University of Dayton Research Institute

The training applications of aircrew performance measurement are numerous. Valid and reliable measures are needed for such purposes as (1) performance feedback to the instructor and the student, (2) student performance diagnosis, (3) student performance assessment and grading, (4) training system evaluation and management, and (5) prediction of job performance. Different types of performance measures are often necessary depending upon the proposed use of the measures in the training environment. The development of adequate performance measures is also required for research and development applications. For example, sensitive measures are needed to evaluate the effectiveness of new training methods and training media such as simulators and part-task trainers.

The panel members discuss why performance measurement development has lagged behind the technology available for recording, analyzing, and displaying performance data. Fundamental problems in measurement development include: the lack of a general theory of human performance, establishing criteria and standards for performance, determining the validity of performance measures, and the fact that important components of performance are not observable. Discussion includes lessons learned in performance measurement development research, techniques and methods for establishing reliability, validity, and useability of measures, and the role of the human observer in performance assessment. Panel members also present the results of ongoing performance measurement research and guidelines for future performance measurement efforts.

Panel

Cardiovascular Reactivity in a Military & U.S. Government Civilian Population

Robert D. Hindelang
Tommie G. Cayton
National Defense University

Suzin E. Mayerson
Uniform Service's University of Health Sciences

Two hundred fifty-one male subjects (207 military officers and 44 government civilians) were administered cardiovascular reactivity testing as part of a health risk appraisal. Reactivity testing consisted of mental arithmetic and video game challenges during which blood pressure and heart rate were recorded at one minute intervals. Thirty-five percent of the total population exceeded the clinical criterion, mean arterial blood pressure of 107. Officers were less reactive than civilians; however, there were no differences among military services. Separate analysis of task challenges indicated officers were significantly less reactive on mental arithmetic. These groups did not differ on resting blood pressure but civilians had greater resting heart rate and were also significantly older. After adjusting for the effects of age and heart rate, no group reactivity differences were found. These results will be discussed in light of possible effects of military attributes as well as measurement issues of cardiovascular reactivity.

Military Officers in Human Factors Engineering:
Responsibilities, Roles, and Realities

In December 1988, DOD Directive 5000.3, Manpower, Personnel, Training and Safety in the Defense Acquisition Process, was approved, mandating that all DOD components in cooperation with industry adhere to human-system integration objectives. 5000.3 states: "The DOD shall maximize the operational effectiveness of all systems . . . by ensuring those systems can be effectively operated, maintained, and supported by well qualified and trained people. To do so, human capabilities and limitations must be fully considered early in the systems design process". Key resources within DOD for supporting human factors research and implementing human engineering in systems design and acquisition are its behavioral scientists and psychologists in uniform. The purpose of this panel is to take a Tri-Service look at what it means to be a military officer involved in the specialized field of human factors engineering. Issues to be treated include: whether there are or should be well-defined career paths for military human factors engineers; unique opportunities, as well as drawbacks and limitations, associated with being a military human factors engineer, and typical job descriptions, compared with the appropriate/ideal utilization of these personnel. The panel will conclude by assessing the current impact of human factors engineering on the acquisition and design process, and by addressing the controversial question of whether or not the DOD even needs military representation among its human factors engineers.

Proposed Panel Members

Moderator: Lieutenant Kristen Morton, USAF
Harry G. Armstrong Aerospace Medical Research Laboratory
Wright-Patterson AFB, OH 45433-6573
AV 986-7142

**Lieutenant Colonel Gerald P. Krueger, USA
US Army Aeromedical Research Laboratory
Fort Rucker, AL 36362-5292
AV 558-6866

Lieutenant Colonel William P. Marshak, USAF
Harry G. Armstrong Aerospace Medical Research Laboratory
Wright-Patterson AFB, OH 45433-6573
AV 785-7555

Lieutenant Colonel Thomas McCloy, USAF
Department of Behavioral Sciences and Leadership
US Air Force Academy, CO 80840-5701
AV 259-4305

Lieutenant Commander Thomas M. Mitchell, USN
Naval Postgraduate School
Monterey, CA 93943
AV 878-2620

Captain Jack D. Mohney, USAF
Aeronautical Systems Division (ASD/ALH)
Wright-Patterson AFB, OH 45433-6503
AV 785-9750

A Structural Assessment of Classroom Learning

Timothy E. Goldsmith, Ph.D.
University of New Mexico

Abstract

This paper describes a technique for assessing classroom learning based on a student's judgments of the interrelationships among the central concepts of a domain. These judgments are subjected to a scaling technique (Pathfinder) to produce a conceptual network. The network is hypothesized to reflect important characteristics of what an individual knows about the domain. This hypothesis was supported by demonstrating that network structures of better students are indeed more similar to an idealized referent structure of the domain than poorer students.

To be knowledgeable in some area is to understand the interrelationships among the important concepts in that domain. It is a long standing assumption within cognitive and educational psychology that this organizational property of knowledge can best be captured with structural representations (e.g., Bower, 1972; Collins & Quillian, 1969). A fundamental problem with conventional educational assessment procedures is that they are in principle, incapable of explicitly representing this configural property of knowledge. Rather, classroom learning is assessed in terms of percentage correct, or in the case of standardized tests, a percentile ranking within some designated population. In some instances performance may be analyzed into subscales, but for the most part learning is represented in terms of a unidimensional scale reflecting the student's relative class standing.

The need for procedures that assess and represent conceptual properties of classroom learning was recognized as early as 1972 by Shavelson and his colleagues (Shavelson, 1972; 1974) who took the view that classroom instruction is most properly seen as the communication of a specific structure that is implicit within the curriculum of a subject matter. Shavelson's work provided the impetus for a slowly developing literature investigating the relationship between derived cognitive structures and domain performance (e.g., Naveh-Benjamin, McKeachie, Lin, & Tucker, 1986). Unfortunately, most of this work has suffered from one or more problems. First, the data on which the cognitive structures were derived were averaged across subjects and therefore failed to assess individual students. If this approach is eventually to be used in the classroom it must be applicable to individual students. Second, the assessment procedures often required students to directly report the organization of their structure. There are reasonable concerns (e.g., Nisbett & Wilson, 1977) as to whether we have direct conscious access to cognitive structure. Third, the derived structures were often assumed to be hierarchical. Hierarchical structures may be appropriate for some domains, but not all. It would be preferable not to constrain the solution to a hierarchical representation. Finally, the basis for comparing the similarity of structures was often subjective. Ideally, we would employ an objective quantitative means of assessing the similarity of representations.

Of the many behavioral changes that may occur with the acquisition of expertise, we are particularly interested in judgments of relatedness among central concepts within the relevant domain. This set of relatedness judgments is assumed to reflect the state of an individual's knowledge. This assumption is tested by examining the predictive validity of such relatedness judgments. However, rather than attempt prediction with the raw ratings, the history of psychological scaling suggests that transformations of the ratings will prove more useful. Therefore, a major effort of our work in classroom assessment has been to determine which transformations of concept ratings have greatest predictive validity. The preference of one transformation over another is determined first by its validity as a predictor of domain performance and second by its representational simplicity relative to the complete data. In the present study we are specifically interested in determining whether Pathfinder (Schvaneveldt, Durso, & Dearholt, 1985) networks have greater predictive validity than raw relatedness ratings or multidimensional scaling (MDS) derived representations.

To assess the predictivenss of a student's cognitive structure we need some means of objectively comparing a student's representation with some idealized referent representation, which in the present work is obtained from the instructor. In the case of raw relatedness ratings, we may simply calculate the product-moment correlation (this and all subsequent correlations are Pearson correlations) between corresponding values in the student's and instructor's ratings. For MDS representations, we first calculate the euclidean distance between all pairs of points in the n-dimensional space and then calculate the correlation between corresponding distances in different representations. A similar approach is used to compare Pathfinder networks but now graph-theoretic distances between nodes in the derived networks are correlated. A second similarity measure for networks is a set-theoretic index of closeness (C) described by Goldsmith and Davenport (in press). Briefly, C examines the degree to which the same node in two graphs is surrounded by a similar neighborhood of nodes. The values of C range between zero (for complementary graphs) to one (for isomorphic graphs). C is assumed to assess structural (i.e., configural) similarity of graphs, and therefore we hypothesize that C applied to Pathfinder networks will indeed allow higher predictive validity of student knowledge than other representation/comparison methods.

The basic purpose of the study was to assess whether the degree of agreement between a student's structure and the instructor's structure was indicative of classroom performance as measured by conventional testing techniques. We hypothesized that students whose structures more closely match the instructor's will indeed be more knowledgeable and hence perform better on standard examinations. We further hypothesized that a measure of representational similarity that assesses configural relationships of knowledge elements will be more predictive of performance than one not based on configural information.

Method

Subjects and Domain

A total of 40 students from a 16-week sophomore/junior level college course on psychological research techniques participated in the study with 20 students coming from each of two separate courses taught in different semesters by the same instructor. A set of 30 central concepts in statistics and experimental design was selected by the instructor in conjunction with other faculty members who taught related courses. Student performance in the course was measured by three exams and two papers totaling 400 points.

Procedure

The purpose of the concept rating project was explained to students at the beginning of the semester. They were told they would rate the relatedness of 435 (30*(30-1)/2) pairs of concepts at the end of the semester, and that these ratings would be used to assess their knowledge of the course material. Students were allowed to earn up to 20 extra course points by performing well on the task. Students performed the task individually and at their own convenience on microcomputers located around campus. The course instructor also rated the concepts to provide a referent structure.

Students were asked to judge the relatedness of each pair of concepts using a 7-point scale where "1" corresponded to less related and "7" to more related. Students were first shown the complete set of concepts and encouraged to pick out concept pairs that were highly related or highly unrelated to serve as anchors. Each pair of concepts appeared left-right centered below the rating scale. A bar marker appeared initially at rating "4" for each concept pair. The bar marker could then be moved with the left and right directional keys on the computer keyboard to select the desired rating. Pressing the space bar accepted the rating and presented the next pair of concepts. The presentation order of the concept pairs was randomized individually for each subject. Additionally, the left-right order of the concepts was randomized for each pair. On average, students took about one hour to complete the set of ratings.

Results

Data from both classes were combined and analyzed. The concept ratings were first converted to proximities by subtracting each rating from eight. These proximities were then analyzed by Kruskal's (1964) nonmetric MDS procedure and Pathfinder. In the case of MDS, an elbow criterion test yielded four dimensions as optimal and so all subsequent MDS analyses are based on four dimensions. Once the MDS and Pathfinder representations were obtained, the similarity between each student's representation and the instructor's was determined using the methods described previously. Comparisons were made from each student's data using each of four different knowledge index variables: correlations on raw ratings, correlations on MDS distances, correlations on Pathfinder graph-theoretic distances, and Pathfinder networks assessed by C. To simplify reporting of the results, we abbreviate these as RAT, MDS, PFR, and PFC, respectively.

Agreement between each student and the instructor was computed based on the four knowledge index variables. Correlations were then computed between each knowledge index variable and the student's earned course points at the end of the semester. The resulting correlations, transformed into r-squared values, for RAT, MDS, PFR, and PFC were .37 .29, .43, and .55 respectively. The correlations were all significant ($p<.01$). The correlation between a student's and instructor's concept ratings (RAT) accounted for 37% of the variance associated with the student's final course grade. Hence, concept ratings themselves are a valid indicator of a student's knowledge. Of more interest, however, is whether scaling algorithms such as MDS and Pathfinder extract from these ratings information that allows even better predictions. The answer varies. Distances from MDS were slightly poorer than ratings in predicting performance whereas Pathfinder distances were better than the ratings. Comparison of Pathfinder networks with C (PFC) provided even better predictions than with correlations (PFR).

A closer look at the relative contribution of each knowledge index for predicting final course points comes from examining partial correlations. Table 1 gives the correlation between each index and final points with the variance contributed by each other index held constant. Consider first Pathfinder networks that have been compared using C. PFC correlates significantly with final points even when each of the other variables is held constant. However, none of the other indices correlates significantly with course grades when the variance contributed by PFC is held constant. This pattern of findings strongly suggests that Pathfinder networks, as assessed by C, are uniquely capturing important predictive variance in the concept ratings. Consider next Pathfinder networks as assessed by correlations. PFR is a significant predictor when ratings and MDS are each held constant, but not PFC.

Table 1
Partial Correlations Between Knowledge Index 1 and Final Course Points with Knowledge Index 2 Held Constant

		Knowledge Index 2			
		RAT	MDS	PFR	PFC
Knowledge Index 1	RAT	-	.34*	.30	.15
	MDS	.03	-	.29	.12
	PFR	.43**	.52**	-	.17
	PFC	.54**	.61**	.46**	-

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

These results imply that Pathfinder networks do indeed contain unique predictive variance over the ratings and MDS, and that a configural assessment of networks is a better index for assessing network similarity than correlations. Apparently, C better reflects commonalities between structures that happen to be important in assessing knowledge. We assume that the characteristics common to a student's structure and instructor's structure that are predictive of knowledge attainment exist at a global or configural level within those representations. This, of course, is exactly the type of information that C is assumed to be good at assessing.

Consider next the results from MDS. Spatial structures did not significantly predict course performance when the other variables were partialled out. MDS has been successful in previous applications for representing physical or continuous relations. Our results may indicate a specific limitation of MDS for assessing conceptual-level relations. Some corroboration for this conclusion is found in work by Cooke, Durso, and Schvaneveldt (1986) who compared MDS and Pathfinder in representing recall data.

Discussion

We conclude that knowledge representations based on college students' concept ratings do indeed offer a valid assessment of their classroom learning. The extent to which a student's representation matches that of the instructor's at the end of the semester is a good index of how much knowledge the student has learned about the domain of study. Further, the predictive ability depends on how the structure is represented and compared.

The cognitive structural approach to classroom assessment has an obvious connection with much of the expert-novice research. Both approaches take a structural representational view of domain-specific knowledge. However, one interesting difference is that the expert-novice work, as the descriptor denotes, compares groups that differ widely in skill levels, whereas the present work attempts to discriminate individuals at a more homogeneous level. Our results show that it is possible to discriminate among students within a common level of expertise and describe these finer-grain differences of performance along a quantitative continuum (i.e., similarity to the instructor's representation). Although this is not to be interpreted as suggesting that qualitative distinctions are unnecessary, it does introduce the possibility that at some levels of analysis expert-novice distinctions can be seen as a continuous transition. More relevant to the problem at hand, the present findings suggest that relatively small differences in classroom knowledge can be discriminated with a cognitive representational approach.

References

Bower, G. H. (1972). A selective review of organizational factors in memory. In E. Tulving & W. Donaldson (Eds.), Organization of Memory (pp. 93-137). New York: Academic Press.

Collins, A. M. & Quillian, M. R. (1969). Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior, 8, 240-247.

Cooke, N. M., Durso, F. T., & Schvaneveldt, R. W. (1986). Recall and measures of memory organization. Journal of Experimental Psychology: Learning, Memory & Cognition, 12, 538-549.

Goldsmith, T. E., & Davenport, D. M. (in press). Assessing structural similarity of graphs. In R. W. Schvaneveldt (Ed.). Pathfinder Associative Networks: Studies in Knowledge Organization, Norwood, NJ: Ablex.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, 29, 115-129.

Naveh-Benjamin, M., McKeachie. W. J., Lin, Y., & Tucker, D. G. (1986). Inferring students' cognitive structures and their development using the "ordered tree technique". Journal of Educational Psychology, 78, 130-140.

Nisbett, R. E. & Wilson, T. D. (1977). Telling more than we know: Verbal reports on mental processes. Psychological Review, 84, 231-259.

Schvaneveldt, R. W., Durso, F. T., & Dearholt, D. W. (1985). Pathfinder: Scaling with Network Structures. Memorandum in Computer and Cognitive Science, MCCS-85-9, Computing Research Laboratory, New Mexico State University.

Shavelson, R. J. (1972). Some aspects of the correspondence between content structure and cognitive structure in physics instruction. Journal of Educational Psychology, 63, 225-234.

Shavelson, R. J. (1974). Methods for examining representations of subject-matter structure in a student's memory. Journal of Research for Science and Teaching, 11, 231-249.

A LEADERSHIP DEVELOPMENT PROGRAM FOR JUNIOR OFFICERS IN STRATEGIC AIR COMMAND

Lt Colonel Kevin W. O'Callaghan, Ph.D.
530 Combat Crew Training Squadron
Plattsburgh Air Force Base, New York

ABSTRACT

Recent changes to the Professional Military Education (PME) development for Air Force officers has led to the formation of a new leadership development initiative by Strategic Air Command. The Lieutenants Professional Development Program (LPDP) was started by SAC to bridge the gap between the lieutenants pre-commissioning training and attendance at Squadron Officers School.

Much has been written over the years about the concept of leadership. According to Air Force Pamphlet 35-49 (1985), "Leadership is the art of influencing and directing people to accomplish the mission. The basic concept the effective leader must keep in mind encompasses two fundamental elements; the mission and the people." For officers in the Air Force, the development of leadership begins during one's pre-commissioning program (ie. ROTC, OTS, AFA), and ultimately results with a commission as a Second Lieutenant. Once on active duty as a lieutenant, it was the usual practice to begin taking your first Professional Military Education Course (Squadron Officer School) by correspondence. However, with recent changes to the Air Force's Professional Military Education, newly commissioned officers do not receive any further formal leadership development training until they reach the rank of captain. Recognizing the gap between an officer's pre-commissioning and eligibility to attend Squadron Officers School, the Commander in Chief of Strategic Air Command (CINSAC) approved the establishment of the Lieutenants Professional Development Program (LPDP).

BACKGROUND

The original LPDP began in September 1979 as a test program. The Commander in Chief of Air Training Command (ATC), concerned about the significant number of lieutenants in key leadership and supervisory positions without any practical experience or specialized management training, tasked the Leadership and Management Development Center (LMDC), to develop and provide a pilot program for selected ATC training centers. The rational for the program was that new officers (particularly from ROTC and OTS) graduating from largely academic oriented programs lacked practical leadership and officership training. The basic focus of the LPDP was to translate leadership and management theory into a program where lieutenants could practice a "hands-on" employment of the leadership principles. From late 1979 until May 1986, LMDC Management Consultant Teams conducted the program at the request of interested organizations.

SAC's Lieutenants Professional Development Program

As stated earlier, with the change to the Air Force PME policy, the CINSAC approved the establishment of the LPDP for all SAC lieutenants on 25 July 1988. The programs objectives were: (1) to strengthen an officers professional values, (2) to provide a foundation for further officer professional development and (3) to translate leadership/management theory into practical application to better prepare officers to assume their responsibilities. To accomplish these objectives, the LPDP consists of a five day, 35-hour program. A description of the LPDP agenda items are found in figure 1. The initial 15 hours (two days) is contracted out to a civilian university (Eastern Washington State University) and is taught by their faculty. The primary aim is for the lieutenants to be exposed to basic management and organizational theory. Some of the topics discussed are: characteristics of organizations and leadership, decision-making, motivation, stress management, human communication and facilitating task groups. Various classroom simulations and exercises are incorporated into the two days to enhance student learning. Recent studies (O'Callaghan,1989) have shown that this approach has proved to be a very effective tool in converting management theory to real world situations routinely faced by lieutenants.

The two day workshop conducted by the university instructors is followed by a three day, 20-hour block of instruction facilitated by a senior Air Force officer. Here the primary focus is on leadership and officership. According to Vice Admiral James B. Stockdale, United States Navy (Ret.) (1982) the concept and importance of leadership is critical to an officer. He states:

> In all that I have been saying, I've made the points that leaders under pressure must keep themselves absolutely clear morally. They must lead by example, must be able to implant high-mindedness in their followers, must have competence beyond status, and must have earned their followers respect by demonstrating integrity.

In an effort to address these and other leadership issues, a variety of topics are discussed during the 20 hours of instruction. These include: the elements of officership; the impact, importance and development of individual values; the importance of ethics; situational leadership; and small group discussions with the Wing Commander, senior officers, senior NCO's and officers from the sister services on pertinent leadership issues. The goal of these discussions is to enable our junior officers to think about and act on their own leadership styles. Concerns about officers not being able to think beyond the scope of black and white issues has been evidenced in a research study conducted by Bryant, O'Callaghan, Buron and Lorenzen (1988). The SAC LPDP is designed to help officers think in the abstract (grey areas) and to better deal with a variety of decisions required as supervisors and managers.

Summary

The LPDP is now conducted at all SAC bases. Its goal is to have all lieutenants assigned to SAC to complete this course prior to being promoted to Captain. The critiques from the participants that have completed the LPDP have been extremely positive. For many of them, the importance of the program can be summed up in the words of one lieutenant's comments on the course critique. When asked what the most beneficitcal aspect of the program was to them, the

FIGURE 1

Description of LPDP Agenda Items

Workshop Introduction

Administrative details for Eastern Washington State University are accomplished during this time, followed by an overview of the LPDP.

Getting Acquainted Exercise

This exercise consists of each workshop participant telling everyone else in the group what their most embarrassing moment was in ROTC, OTS, AFA, Charm School, or active duty; three adjectives that describe themselves; the finest leader they know and; their symbol for leadership. This exercise is designed to both "break and ice" and to allow the members of the group to know a little more about each other.

Transition on being a Manager

This lecture and discussion describes the new roles as managers lieutenants take on as a result of their commission into the Air Force. Factors they must face, planning and organzing issues are looked at.

Setting Goals

This session deals with the importance of goal setting and maintaining standards for running an effective organization. Problems faced by not taking time to set goals will also be discussed.

Characteristics of Organizations

This lecture and discussion will focus on the importance roles and norms play in organizations. Typical behaviors that characterize a person in a specific social/work context as well as a look at the generally agreed upon informal rules that govern and guide group members behavior will be discussed.

Communications

This exercise, known as "the Domino Game," is designed to test the ability of the workshop participants, to communicate complex information. Most participants find it fun and exciting and will continue the game as long as allowed.

Organizational Theory

This lecture will expose participants to a sampling of motivational theories and applied to such areas as performance, satisfaction, leadership and organizational climate.

Time Management

During this block of instruction, the importance of proper time management skills will be discussed. Ways of better utilizing one's time, money and other resources will be looked at.

Motivation and Punishment

This lecture covers rational approaches to goal setting, delegation, the five core job dimesions to work satisfaction, power and emotional factors in motivating others.

Positive Motivation Model

This lecture will take a look at the Positive Motivation Model. This is a five step model that will introduce participants to the importance of expectations, skills, feedback, consequences and growth.

Problem Solving Exercise

This block of instruction uses the "Desert Survival" exercise. Participants are given a scenario and told to rank order solutions individually then later as a group. The exercise enables participants to understand the basic components of problem solving and differences between working as an individual vs a group.

Tower Building Exercise

This activity shows how well the group plans, delegates, uses resources, communicates and works together. In the exercise, each group is given a set of tinker toys and are told they have 20 minutes to plan and practice building the tallest freestanding structure possible. However, they cannot put any of the pieces together during the 20 minute planning phase; instead at the end of this phase, each group has only 40 seconds to actually build the structure.

Verbal/Nonverble Communication/Hostile Groups

This block ties together all that has been looked at the past two days of the workshop with the emphasis on communications. In addition, looking at how to manage a hostile audience or group will also be discussed.

Stress Management

This very important block will focus on the importance of stress management in being an effective officer. Symptoms and causes of stress, potential issues and concerns with stress and possible consequences of stress will be looked at. In addition various techniques for dealing with and coping with stress will be presented.

Introduction to Military Portion

This will begin the three day Air Force portion of the LPDP. It will consist of administrative policies and procedures and student requirements, followed by the welcome by the Wing Commander. The purpose and background of the program will be explained.

Officership/Standards

Key elements of officership; the importance of officership and Air Force standards to effective leadership; discussion on enforcing standards through the use of exercises will be used.

Values/Ethics

During this block, the significance of how we develop our value system and how it relates to individuals and groups will be discussed. A number of exercises on values/ethics are incorporated as well as a PBS video an ethics.

Leadership

Situational leadership is the focus of this block as participants analyze this theory using the film "Twelve O'Clock High." Also used is another film, "A Leadership Alliance" by Tom Peters to illustrate different leadership approaches to delegation.

Military Justice

A briefing by the SJA introduces professional relationships with subordinates, peers, and supervisors. Possible consequences of unprofessional relationships are discussed as it relates to officership.

Professional Relationships

This phase of the workshops gives the participants an opportunity to have a personal interview/discussion with senior officers, senior NCO's and officers from the joint service. A question and answer period, conducted in an informal forum enables discussion of their views of officers, professional relations and any topic related to this program.

POW Video

Colonel Henry Fowler, Air Force SJA, talks about his life as a POW for almost six years, and the impact of leadership in a POW environment.

End of Course Critique

LPDP wrap up items will be discussed followed by an end-of-course student critique.

lieutenant responded: "Officership. I have had a lot of experience in the leadership and management area, but I have not had any training in officership. I found after the exercises and discussions I knew what it was but I hadn't identified it as such. It was a tremendous experience for me." Many of the comments gathered from the end of course critiques echos this lieutenant's thoughts. It appears evident that the importance of PME is very relevant to the development of our future Air Force officers and leaders. The Lieutenants Professional Development Program is Strategic Air Command's attempt at providing its junior officers the opportunity to ask questions, give their opinions, make decisions and develop and practice their own leadership styles. If the comments gathered thus far are any indication of the success of the LPDP, the program will be a valuable asset to the Air Force.

References

Air Force Pamphlet 35-49 (1985). Air Force Leadership. Department of the Air Force, Headquarters United States Air Force, Washington, D.C.

(Anonymous, 1988). Lieutenants Professional Development Program Curriculum Guide. Headquarters Strategic Air Command, Offutt Air Force Base, Nebraska.

Bryant,C., O'Callaghan, K.W., Buron, R.J., Lorenzen, G.L., (1988). The cognitive development of Air Force officer candidates. Psychology in the Department of Defense, 1, 414-417.

O'Callaghan, K.W., (1989). An effective leadership education program for junior level executives and managers. Presentation at the Eastern Washington State University Instructors Conference. Las Vegas, Nevada.

Stockdale, J.B. (1982). Machiavelli, management, and moral leadership. U.S. Air Force Academy Journal of Professional Military Ethics, August.

Competition, Cooperation and the Performance of Complex Tasks

David B. Porter, D.Phil
Megan E. Bird
Arthur Wunder
US Air Force Academy

Forty eight Air Force Academy cadets were used to examine the effects of cooperative vs competitive attitudes and reward structures on the performance of an arcade-type computer game. Although neither factor alone influenced performance, their interaction was found to be very significant; subjects performed poorly when the reward structure did not match their individual orientation. Implications for military training programs are discussed.

Morton Deutsch (1949) identified two types of situations as end points on a continuum of outcome relationships. He defined cooperative situations as those in which the goals of all individuals were positively related (i.e., the success of any member of a group made the success of other members more likely). In contrast, competitive situations were defined as those in which the success of one member made the success of others less probable. Deutsch found that attitudes, perception, and performance can be influenced by these different reward structures. Since Deutsch, educational, social and industrial psychologists have studied the effects of these alternative reward structures on a variety of tasks in many different settings. Although most published reports suggest advantages for cooperative reward structures (Johnson and Johnson, 1983), many institutions continue to attempt to use individual competition to increase motivation and enhance performance.

Individuals differ in their preference for these reward structures and these preferences affect a variety of outcomes (Cohen, 1982). Preferences reflect underlying belief structures and values. These "personal epistemologies" can be more important predictors of performance and satisfaction than traditional individual difference measures such as "speed" or "intelligence' (Greeno, 1989). Most studies of individual differences support the notion that cooperative individuals are more likely to be associated with higher group satisfaction.

In this study, we sought to vary these two factors independently: the reward structure and individual attitudes toward these structures. The reward structure was manipulated to create either competitive or cooperative interpersonal relations within small groups. Based on subjects' responses to an attitude survey, they were selectively assigned to create groups of individuals with either competitive or cooperative orientations. The primary criterion was performance on a demanding, arcade-type computer game modified to require coordinated inputs from two subjects. Specifically, this study sought to identify the main and interactive effects of situational and individual factors on performance and satisfaction. It was hypothesized that advantages associated with cooperative reward structures and cooperative individuals suggested by educational and industrial research would extend to this complex interactive task.

Method

Forty-eight freshmen cadets (forty-one male and seven female) between the ages of 18 and 21 volunteered to participate in this study. Although subjects did not receive remuneration, participation in this experiment excused them from two hours of unpleasant training and prizes were offered for top performers. As cadets, our subjects were part of a select group of young adults. Cadets average SAT score is 1230, 73% were in the top 10% of their graduating class, and 80% earned one or more varsity letters in high school athletics. As a group, cadets display such homogeneity on many measures of academic and athletic achievement, they are a poor experimental sample. Their distinctiveness from the general population and restricted range are both problematic. However, cadets do represent many young adults who are competitively selected for entry and education in demanding military and civilian professions. It is also true that despite their similarity in achievements, there is considerable diversity in some cadet attitudes.

A 10-item attitudinal survey was used to measure subjects' preferences for cooperative or competitive environments. Subjects were required to indicate their agreement or disagreement with each of 10 statements on a 6-point scale. Five of the statements reflected individually competitive orientations (e.g., "I prefer individual sports to team sports") and five statements reflected cooperative orientations (e.g., "I take pride in the accomplishments of others"). The sum of a subject's responses to the five competitive statements were subtracted from the sum of their responses to the cooperative statements. The difference was used as a measure of cooperative orientation.

The Whale Game, an arcade-type computer game was the criterion task. Specially developed to study human performance, this game required subjects to guide a blue whale to accomplish two distinctly different tasks: eating plankton and crashing kayaks. Plankton eating was a tracking task which required subjects to guide the whale to a randomly drifting plankton target. This task was simple but the unpredictable movement of the plankton made it very uncertain. In contrast, kayak crashing was certain but complex. Twenty-five kayaks appeared at discrete locations on the screen perimeter at predetermined times throughout the game. Once on screen, the kayaks moved one space horizontally and vertically toward the whale. This pursuit continued until one of two things happened: the kayak reached the whale or it crashed into one of twenty icebergs. Subjects lost points for being harpooned but gained points if the kayaks crashed. Subjects were given explicit priority instructions at the beginning of each game telling them which of the tasks was more important (i.e., worth 10 times more points).

Groups of eight subjects completed the attitudinal survey and received basic instruction on the Whale Game during the initial one-hour session. They accomplished 8 practice trials (4 of each priority) and recorded their scores at the end of each trial. Subjects were then assigned to eight six-person groups such that each group's average performance during the initial session was equivalent. Seven of the groups contained one female and five male cadets and the other was all male. No group had more than two members from the same squadron. Within these constraints, distinctly competitive or cooperative groups were created. This treatment was designed to accentuate the effects of individual attitudes.

The one-hour experimental session took place approximately six weeks after the initial training session. The second independent variable in this study was the task reward structure. Four groups received competitive instructions and the other four received cooperative instructions. A competitive reward structure was imposed by telling the six subjects in a particular group that the top scorer among them would be invited to a "pizza party". Although most college freshmen would not see this reward as being very significant, our subjects were quite excited by it. The same reward was offered to the cooperative groups but the criterion was changed. The cooperative reward structure was established by telling subjects that the score of all group members would be compared to the scores of other groups and the group with the combined top score would be invited together to the pizza party. Thus in the competitive groups, the object was to perform better than the other members. In the cooperative groups, the goal was to maximize the performance of the whole group.

Two types of dependent measures were of interest. Task performance was measured by the total number of points scored on the two game tasks: eating plankton and crashing kayaks. The game itself was identical to the one subjects practiced during the initial session but the way it was played was not. The six subjects in each group shared three computers. For each trial, one subject controlled the horizontal movement of the whale and another controlled the vertical movement. This manipulation forced subjects to interact with one another (as would an aircrew). Subjects rotated so they played one game of each priority with each of the other five subjects. Each subject controlled horizontal whale movement during five trials and vertical movement on the other five. Both subjects shared the same score for a given trial but the rotation allowed overall individual scores to vary greatly. At the end of the session, subjects were asked to respond to five questions using a six point Lickert scale. The questions concerned their general satisfaction with the game and their willingness to participate in future experiments. The average rating on these five questions was used as a measure of their satisfaction, the second dependent variable.

Results

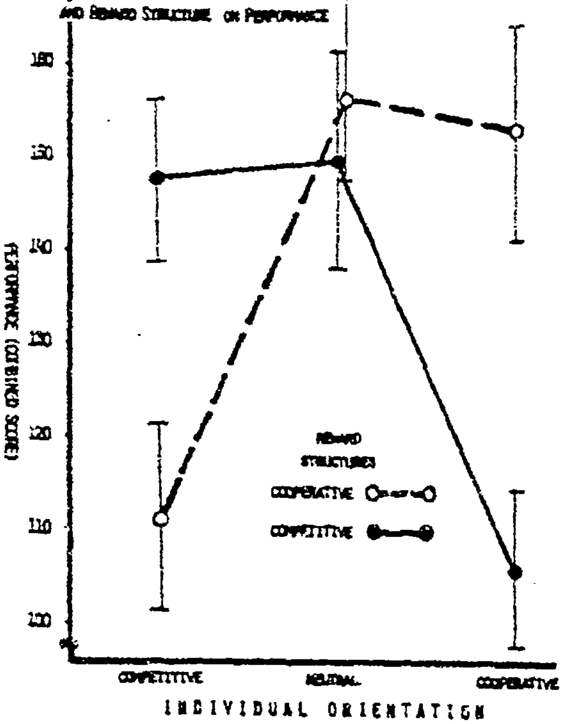
Results from the attitudinal questions showed that our subjects tended to endorse the cooperative statements slightly more strongly than the competitive ones. The mean was 2.52 (s=4.27) and individual scores ranged from -6 to 11. The distribution appeared to be normal.

Although data could have been partitioned to examine the effects of practice or differences in the two subtasks, the power of this design only supported using a single aggregate measure of performance for each subject. Multiple regression analysis was used to identify the effects of attitudes and reward structure on this single measure of overall performance. In the first step of the analysis, both attitude and reward structure were entered into the regression equation. The main effects of these two variables explained less than 1.0 % of the total variance in performance. However, when the interaction (reflected by the product of the two variables) was entered, the portion of the variance explained (r^2) increased to 20.7 %. The beta regression equation listed below was significant at the .05 level ($F(3,44)=3.819$).

Score = -.56 Attitude -.23 Rewards +.71 Interaction

To depict these results graphically, subjects were divided into three groups (competitive, neutral and cooperative) and their respective mean scores (and standard errors) were plotted separately for the two reward structures. The interaction can be clearly seen in Figure 1 immediately to the right. Clearly the greatest effects of reward structure were seen in the performance of those with the most pronounced attitudes toward either competition or cooperation.

FIGURE 1. The Effects of Individual Orientation and Reward Structure on Performance



Subject satisfaction was also an important dependent measure. In this case, however, the analysis showed that only subjects' attitudes had a significant effect. The correlation between attitude and satisfaction was .354 ($df=46$, $p<.01$, two tailed). Cooperative individuals were more satisfied and willing to participate in future experiments. This result is not too surprising; what is surprising is the lack of significant relationships between satisfaction and reward structure ($r=.058$), the reward structure interaction ($r=.150$) and even task performance ($r=.041$).

Briefly, these results suggest performance was better when the reward structure matched the individual's orientation. This effect appeared to be particularly detrimental to those individuals with strong attitudes who found themselves under "inappropriate" reward structures. Those with moderate attitudes appeared to be able to perform well in either environment. Satisfaction was unaffected by performance or reward structure but was generally higher among those who expressed a preference for cooperation.

Discussion

The expected direct advantages of cooperative attitudes and reward structures were not found. However, the pattern of results still provides interesting insights into the dynamics which may underlie other successful demonstrations of the benefits of cooperation.

Earlier the issue was raised that Air Force Academy cadets differ somewhat from other college freshmen. One difference is the high level and variety of their high school achievements. If one assumes that a competitive attitude is positively associated with the academic, athletic and extracurricular achievements used for selection, our sample may have been more competitive than subjects in other studies of cooperative effects. If only those subjects with relatively cooperative attitudes are considered, there appears to be clear support for the proposition that cooperative reward structures increase performance.

Although attempts were made to employ an experimental task which was more ecologically valid, the duration of the experiment was still less than one hour. Most studies which have successfully demonstrated the advantages of cooperation have involved weeks, months or even years. Might we expect advantages to accrue to the cooperative groups if this experiment were extended? Cooperative individuals were more satisfied and also more willing to participate in future experiments regardless of the reward structure or level of performance. To the extent this increased motivation result in greater persistence and more rapid skill acquisition, the cooperative subjects should continue to improve in comparison to their less satisfied competitive peers. Additionally, it might be expected that positive relationships and social support systems would be more likely to develop among the cooperative groups.

The implications of these results for military training programs are considerable. Many training programs rely heavily on interpersonal competition as a mechanism for increasing trainees' motivation and hence performance. Our results suggest that the short term effect of such a policy on performance is nil. Emphasis on competition will enhance the performance of some trainees but inhibit the performance of others. This begs the question of which trainees are helped and which are hindered by such a policy. Ideally, those who are most likely to become competent and dedicated members of operational units are those whose performance would be facilitated. However, just the opposite result occurs if competitive reward structures are employed to train individuals for operational environments in which teamwork and cooperation are required. Competitive reward structures benefit competitive individuals. Competitive individuals, especially those with pronounced biases, are likely to perform poorly in cooperative environments. Additionally, competitive individuals seem less likely to be satisfied regardless of their level of performance or the type of reward structure. From this perspective, part of the answer to the question "Where have all our pilots gone?" becomes distressingly clear.

References

Deutsch, M.A. (1949). A theory of cooperation and competition. Human Relations, 2(2), 129-152.

Cohen, J. (1982). Cooperative and competitive styles - the construct and its relevance. Human Relations, 35, 621-633.

Greeno, J.G. (1989). A perspective on thinking. American Psychologist, 44(2), 134-141.

Johnson, D.W. & Johnson, R.T. (1983). The socialization and achievement crisis: are cooperative learning experiences the solution? in L. Bickman (Ed.), Applied social psychology annual 4. Beverly Hills: Sage

Authors Notes

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

The authors gratefully acknowledge the support of the Frank J. Seiler Research Laboratory, US Air Force Academy.

Military Career Exploration using the Armed Services Vocational Aptitude Battery and Personal Type Inventory

Kenneth W. Wegner
Boston College

Abstract

This paper describes the development and use of a new Personal Type Inventory and Military Career Options listings for counseling and placing military entrants. Methods and advantages for their use in assisting enlisted (in conjunction with ASVAB scores) and officer personnel in choosing military careers are presented.

Career and occupational psychologists have been familiar for years with Holland's (1985) personality typology. Briefly he proposes that there are six primary personality types (Realistic, Investigative, Artistic, Social, Enterprising, and Conventional), and the individual with a certain personality type tends to seek out and interact "best" with a matching environment. His current work is focused on the Self Directed Search (1985) (SDS), in which the respondent indicates their competencies, activity and occupational preferences, and estimated abilities. These responses are summarized as a three letter occupational code. An occupational classification booklet is provided to locate suitable work environments that match this code.

Several years ago this writer began to wonder why Holland had not yet developed a more generalized personality instrument to define his typology, since he lists a variety of personality descriptions for each of the six types (SDS Professional Manual, 1985). It was decided to attempt to develop a Personal Type Inventory (PTI) using similar descriptors to determine what more generalized personality (as opposed to vocational) typology might emerge. In reviewing Holland's lists, two general handicaps to their use in a self-descriptive inventory became obvious. The first is that the same descriptor is frequently used for more than one type (e.g., sociable describes both the Social and Enterprising types). Secondly, a variety of negative descriptors (e.g. asocial, pessimistic, disorderly, domineering, and unimaginative) are contained in his lists.

PTI Development

Briefly, the approach to developing appropriate items was to eliminate all overlapping and less socially desirable descriptors suggested by Holland, and generate lists of semantically equivalent descriptors for each type. The item format chosen was paired-comparison with each word of a pair representing one of Holland's six types. Holland proposed that the six types form a hexagonal model (in two-dimensional space) with more closely related types adjacent to each other and less closely related types opposite each other.

It appeared logical from this model to pair descriptors that represented non-adjacent types. A second criteria used was to pair descriptors from the most infrequently occurring two letter Holland typology codes. For example, the fewest members of any sample receive an Artistic-Conventional, or Conventional-Artistic classification on the SDS. A primary reason for using these criteria and the paired-comparison item format is that it forces a higher degree of consistency of response in the individual and increases the differentiation of the type scores.

After various item try-outs, 90 paired-comparison items were developed that can be scored for six scales which parallel Holland's themes : Practical (P), Reflective (R), Imaginative (I), Supportive (S), Extroverted (E), and Disciplined (D). The names of the scales were determined from their general item content and a factor analysis of the responses of 376 male and female subjects. Three bi-polar factors emerged from a Harris image analysis oblique solution. Significant factor loadings were found for the following scales: E vs. R and P, S vs. I and D, and I vs. D and R. These analyses also suggest a three dimensional model rather than the two dimensional hexagonal model as proposed by Holland (1985). Other correlational and factor analyses indicate that the scales are also gender-free. Their test-retest reliabilities range from .82 to .90, which compares favorably with other similar personality and interest inventories.

The PTI has now been administered to over 500 college level students. Concurrent validity studies indicate that the PTI will classify students by college major with approximately the same relative efficiencies as Holland's SDS and similar instruments. The primary advantage for the PTI is the 15 to 30 minutes required for administration and self-scoring. Field trials with junior high and high school students are currently underway. A spanish language form has been developed and field-tested with employed subjects in the Guadalajara, Mexico area.

Occupations Lists Development

In order to provide career exploration information, a PTI Occupations List was developed to classify approximately 500 occupations (90%-95% of those in the U.S. labor market) using two letter PTI codes (e.g., Practical-Reflective). The assignment of two letter codes to occupations was determined by utilizing a combination of two well researched systems. The primary system was the 12 interest based categories used in the Guide for Occupational Exploration (1979). This was combined with the orientations of occupations to Data, People, Things and Ideas framework utilized in the development of the Dictionary of Occupational Titles (1977) occupational codes.

At the same time Military Career Options (MCO) listings were developed to relate PTI two-letter codes to the 134 enlisted and 71 officer occupations contained in the 1988-89 Military Career Guide (1988). For enlisted occupations, the major categorization is by the four Armed Forces Vocational Aptitude Battery, Form 14 (ASVAB, 1984) Occupational Groups: (1) Mechanical and Crafts; (2) Business and

Clerical; (3) Electronics and Electrical; and (4) Health, Social, and Technology. The enlisted occupations are sub-categorized by PTI two-letter codes under each of these Occupational Groups. This list also indicates the ASVAB Form 14 "50-50" score (representing the 50 percent probability of qualifying for training) for each enlisted occupation. The 71 Officer occupations were classified by two letter PTI codes. A partial listing for the MCO enlisted category of Business and Clerical is illustrated below.

Business and Clerical

PTI Code	ASVAB 50-50 Score
DE	
Lodging Specialists	26
Postal Specialists	29
Transportation Specialists	24
DS	
Administrative Support Specialists	28
Personnel Specialists	30
Religious Program Specialists	29

Administration and Use

The procedure for using this system with enlisted or officer applicants is to administer the PTI, and then search for occupations matching combinations of the obtained three highest score code letters in the MCO lists. Analysis of highest through lowest obtained PTI scores for 376 subjects indicates that only a letter representing a score of 17 or above should be used as the first letter in a code. Similarly, only letters representing a score of 15 or above should be used as a second letter in a code. Approximately two thirds of examinees will have three scores of 15 or above, and thus would have six two letter code combinations to apply in the MCO lists. Only about two percent would be limited to one PTI code using these rules.

To find possible enlisted career options in the MCO, one would first look for matching PTI code occupations under the occupational group representing their highest ASVAB score (e.g., Business and Clerical). If none of their PTI code combinations appear there, they would move on to their next highest ASVAB occupational group score and repeat the search. Although military occupations have limited variety compared with the civilian labor market, all six major PTI types are represented in both the enlisted and officers occupational lists. However, enlisted military occupations are heavily concentrated in those areas associated with Practical, Reflective, and Disciplined types (approximately 75 percent). This is less true of officer occupations where a more managerial orientation (Extroverted type) is also present in many type codes. After a short list of appropriate occupations has been developed, the individual may be referred to the Military Career Guide (1988) for a description of the characteristics and requirements of those specialties.

Practical Advantages

The use of the PTI with military entrants (particularly in combination with ASVAB scores for enlisted personnel) provides several practical advantages in military career selection. The PTI is self-scoring and requires 15 to 30 minutes to complete. It may also be utilized in individual or group situations. For enlisted examinees it can be administered at the same time as the ASVAB or independently. Extensive training is not needed in administration of these instruments, since the directions for using the PTI and MCO are self-explanatory. However the examinee will obviously need the same guidance in using informational sources about military careers as is needed for counseling related to ASVAB scores.

The PTI may also be useful as an adjunct to (or perhaps suitable substitute for) the Exploring Careers (1988) workbook. Used in conjunction with the workbook, the three highest PTI scores could provide supplementary data for determining the individual's three most relevant areas of interests and skills. For example, a high score on the PTI Supportive scale would suggest the selection of Work With People as a high interest and skill area in the Exploring Careers exercises.

The PTI scores and MCO officer list would also appear to be a most efficient method for making career choices among ROTC, OCS, and military academy students. The writer's experience as a professional counselor and former Air Force Academy and AFROTC Liaison Officer suggest such a system would be quite helpful in counseling applicants. Similarly, it could be used in the final year of military academy or ROTC training to provide guidance in selecting military careers.

References

Holland, J.L. (1985). Professional manual for the Self-Directed Search. Odessa, FL. Psychological Assessment Resources, Inc.

U. S. Department of Defense (1984). Counselor's manual for the Armed Services Vocational Aptitude Battery Form 14. North Chicago, IL: U.S.Military Entrance Processing Command.

U. S. Department of Defense (1988). Exploring Careers: The ASVAB Workbook. Washington, DC: U.S. Government Printing Office.

U. S. Department of Defense (1988). Military Career Guide. (1988-89).Washington, DC: U.S. Government Printing Office.

U. S. Department of Labor (1977). Dictionary of Occupational Titles (4th ed.). Washington, DC: U.S. Government Printing Office.

U. S. Department of Labor (1979). Guide for Occupational Exploration.Washington, DC: U.S. Government Printing Office.

Wegner, K.W. (1990). Personal Type Inventory. Largo, FL: Careers, Inc.

THE EFFECT OF "COSMETIC" DESIGN VARIATION ON PERCEIVED VEHICLE SAFETY

Dennis B. Beringer
New Mexico State University

Abstract

Vehicle systems come in a variety of shapes and sizes. Some of the noticeable external variations are direct reflections of structural differences while others are more "cosmetic" in nature. It was hypothesized that judgements of vehicle safety, although often a function of structural design and operator personal experience, may be influenced by essentially cosmetic external variations in design that do not represent salient differences in vehicle structure. Individuals were asked to rate the general and specific safety characteristics of automobiles from viewing a series of photographs of each of several body styles. Although the general rank ordering placing sedans and station wagons safest with convertibles least safe was expected, models that were structurally equivalent were rated differently depending upon whether a removable roof panel was installed or removed. Thus it was determined that some removable sections or accessories may actually decrease the operator's perception of the hazard present although the actual hazard may not have been decreased.

Examination of contemporary vehicle systems, be they designed to traverse the ground, cross the water or fly through the air, reveals a tremendous diversity of designs. In some cases the variations that one finds within a specific category of vehicle are a direct function of variations in the task to be performed. If one examines the structure of a motorcycle, for example, in comparison with a pick-up truck, these two vehicles are clearly designed to perform different tasks. A byproduct of the differences is the enclosing of the operator in the truck, providing some buffer between the operator and the elements, be they weather or other vehicles. Thus one would expect some difference in the rating of vehicle safety based upon the presence/absence of operator containment (and the stability of the vehicle; number of wheels, center of gravity, etc.).

A rather unique situation develops when a single system may take on several configurations. This is particularly interesting when changes between configurations do not materially affect the structural integrity of the system but may alter the exterior appearance. One primary example of this type of system is the case of vehicles having removable roof sections. In some of these vehicles one may interchange different types of roof sections (tinted glass, fiberglas) or operate the vehicle without any such sections in place. This variation in configuration may play a role in the relative rating of vehicle safety so obtained, particularly if the rating is largely determined with containment/retention of the operator in an accident or rollover as a major factor.

The question of how these perceptions of hazard might vary arose recently in litigation. Misperceptions in automotive operations have been examined in several cases (Vernoy and Tomerlin, 1989; Mac Gregor and Slovic, 1989) as has risk taking and accidents (Taylor, 1976). It is generally accepted that an accurate perception of a hazard is necessary as part of the process of avoiding an accident (Smith and Beringer, 1987). A case had been identified where perception of hazard might not be accurate for part of the user population.

In particular, an automobile was identified (Chevrolet Corvette) that could take several configurations; the T-top design allowed one to configure the vehicle as either a pseudo hard top (plexiglas panels), sun roof (tinted glass panels), or pseudo convertible (panels removed). The T-top configuration consists of two panels, one on either side of the vehicle centerline. The two panels are separated by a central rib. Removal of the panels creates two rectangular openings, one above the driver and one above the passenger. It was generally believed that individuals were likely to perceive the configurations with installed panels as slightly safer than that with the panels removed due to the perception that the panels provided some measure of containment for the operator. In most cases this would be true given that the panels, particularly those made of fiberglas, had a positive retention system that prevented the panels from being released or caused the panels to become an integral part of the structure when fastened in place. This particular system, however, was found to release the panels under certain conditions in a roll-over accident. In one particular case a panel was determined to have been released during the first roll in such an accident.

The result in this particular case was that the probability of being ejected through the roof in equivalent accidents would not be materially affected by the presence or absence of the roof panels because they would be likely to be missing early in the progress of the accident. Thus an unwarranted differential perception of hazard as a function of the presence/absence of roof panels might be a factor in determining whether the operator was likely to wear a seat belt (prior to mandatory seat belt useage laws). While risk taking as function of seat belt useage has been investigated (Evans, Wasielewski, and von Busek, 1982), no study appeared to have examined the reverse where vehicle design was examined as the causal factor in seat belt useage. A pictorial survey was thus constructed to examine the perceptions of potential operators (drivers) of this system as well as other automotive systems likely to be encountered by the user population.

Method

Stimuli

A series of photographs was prepared that represented a wide variety of automobile styles. These included the following models: Jetta convertible, Datsun 280Z, Camaro, Cutlass Salon, compact Chevrolet station wagon, and Chevrolet Corvette. The convertible was photographed both with the top up and the top down. The Corvette was photographed with fiberglas panels in place, tinted glass panels in place, and panels removed. All of the photographed automobiles were white. Each of the automobiles was photographed from the driver's side at three distances; far with the automobile filling approximately 60% of the frame width, intermediate with the automobile filling the width of the frame, and near with the automobile extending beyond the width of the frame. The latter was intended to provide a detailed view of the driver's side as it would be seen upon approach to the automobile as though one was about to enter it. Details of the roofline were thus visible in this close view.

Procedures

Subjects were students in introductory psychology and were measured in a group setting. The experimenters were allowed to show the class 35mm slides of the automobiles at the end of a class session. Each person was given a questionaire form to fill out during the viewing of the slides. The photographs were shown in the order of far to near as though one was walking toward the vehicle. Each slide was shown for approximately 5 seconds with a longer pause between vehicle types (sets of three photos). Subjects rated each vehicle at each portrayed distance using the provided questionaire.

Questionaire. The rating form used had four parts. Part I was titled "General Safety Rating: How safe do you think is each of the following cars?" One scale was used for each of the nine (9) conditions with all nine scales appearing on one page. The seven rating categories were labeled as follows from left to right: very unsafe, moderately unsafe, slightly unsafe, average, slightly safe, moderately safe, very safe. Subejcts were to mark an "A" on the scale for the far view, a "B" for the intermediate view, and a "C" for the near view. The second page of the form was labeled, "Likelihood of INJURY: How likely do you think it is that a person could be seriously injured in this car?" The same number and type of scales were again used but the labels were altered to read: very unlikely, moderately unlikely, slightly unlikely, unsure, slightly likely, moderately likely, very likely.

The third page of the form was titled, "Likelihood of BEING THROWN from CAR: How likely do you think it is that a person could be thrown from this car in a rollover?" The same scales, although physically smaller versions, were used for responses and two columns of these scales were used; one for "wearing a seat belt" and one for "NOT wearing a seat belt". It should be noted that the entire set of slides was shown for each of the questionaire pages. Thus when filling out this third page subjects had already seen each of the stimuli twice and were viewing each a third time. A fourth page was used to assess whether the individual felt it desireable or undesirable to be thrown from a car during an accident, subject age, gender, seat belt use, and whether each was or was not a licensed driver (experience, exposure). Scoring was accomplished by assigning the numbers 1 through 7 to the seven scale categories.

Results

General Safety Rating

The general ordering obtained was expected, placing the automobile types in the following from rated most safe to rated least safe: station wagon, Camaro, Cutlass Salon, 280Z, Corvette (glass), Jetta (top up), Corvette (fiberglas), Corvette (open), and Jetta (top down). An interaction of body style and viewing distance emerged, with the Cutlass and Camaro exchanging ranks as the station wagon and cutlass increased in rated safety while others decreased in rated safety as viewing distance decreased. This results, depicted in Figure 1, is undoubtedly due to the relative size of the cars becoming more evident as view distance was decreased and roof line features, etc., becoming more discernable (sun roof openings in particular). The interesting difference was the general ranking of the open Corvette as more hazardous than the same automobile with either glass or fiberglas panels in place.

Perceived Likelihood of Ejection

Although the ordering of vehicles was essentially the same as before, the seat-belt variable was responsible for a shift of approximately two categories on the scale when likelihood of ejection in a rollover accident was rated. On the top end of the distribution the Jetta convertible (open) moved from a 4.65 (slightly likely) with a belt to 6.55 (very likely) without a belt. The station wagon, at the other extreme, moved from 1.6 (moderately to slightly unlikely) with seatbelt to 3.85 (slightly likely) without.

The conditions of interest were those involving the Corvette configurations. Mean ratings for these three forms of the same automotive system were all within 0.6 points in the belted condition: fiberglas, 2.94; glass, 3.28; open, 3.56 (1 = very unlikely, 7 = very likely). In the no-belt condition, however, the separation increased: fiberglas, 5.06; glass, 5.31; open, 6.22. A Friedman analysis of variance by ranks indicated a significant difference across these ratings [$\chi_r^2(2) = 8.453$, $p<.025$; when corrected for ties, $\chi_r^2>14.0$]. Comparison of the glass and open configurations using the Friedman test indicated near significance (tabled value for p=.05 was 3.84; obtained value was 3.78). The difference between the open top and fiberglas-panel conditions was significant as indicated by the earlier test. Thus these configurations were rated as different in terms of rollover ejection hazard although the actual hazard was likely to be identical. It should also be noted that the ratings of possible serious injury closely followed the ratings of potential ejection from the vehicle.

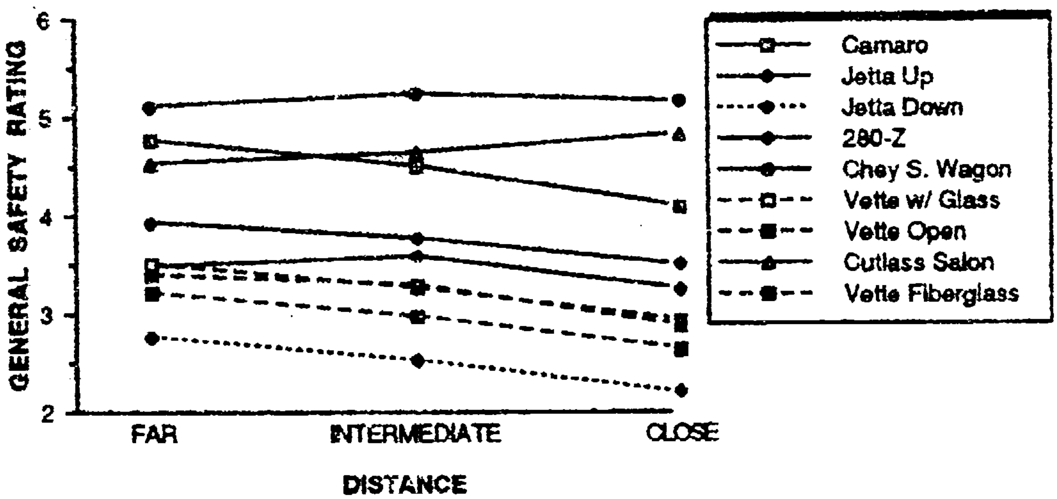


Figure 1. General safety ratings by vehicle type and distance.

Discussion and Conclusions

It is apparent that the perception of hazard in this situation was not equivalent to the true physical hazard as embodied in the design. This is undoubtedly due in large part to the general expectation that parts of a system that appear integral or well attached will remain so in an accident. This is, of course, a tenuous assumption at best given engineering considerations and the forces generated in many automotive accidents. The solution may be one of simply informing the system user of the potential hasard in a salient way. One cannot, however, even begin such a discussion without immediately stumbling upon the effectiveness of warning systems, signs, labels, instructions, and the like. One has the choice, then, of either engineering systems to meet expectations or changing expectations to coincide with actual system performance limitations. Once again it appears, as always, that there are no simple solutions. This issue may become a factor in military systems where modular designs create an appearance of unity in a system but where, in fact, the system may come apart as a result of accidents, hostile action, etc. The likelihood is much lower, however, for military systems than for comparable consumer products not designed for survival in high-stress environments.

References

Evans, L., Wasielewski, P. and von Busek, C. R. (1982). Compulsory seat belt useage and driver risk-taking behavior. Human Factors, 24(1), 41-60.

MacGregor, D. G. and Slovic, P. (1989). Perception of risk in automotive systems. Human Factors, 31(4), 377-390.

Smith, M. J. and Beringer, D. B. (1987). Human factors in occupational injury evaluation and control. Chapter 7.1 in G. Salvendy (Ed.), Handbook of Human Factors, 767-789.

Taylor, D. H. (1976). Accidents, Risks, and models of explanation. Human Factors, 18(4), 371-380.

Vernoy, M. W. and Tomerlin, J. (1989). Pedal error and misperceived centerline in eight different automobiles. Human Factors, 31(4), 369-375.

Human Factors in the Manufacture of Military Uniforms[1]

Michael J. Kelly, Ph.D.
Daniel J. Ortiz
Dennis J. Folds, Ph.D.
Theodore K. Courtney
Georgia Institute of Technology

Abstract

Manufacturing military uniforms is a labor intensive process requiring complex perceptual and motor skills and repetitive manual operations. The purpose of this research was to improve sewing operators' efficiency and productivity by identifying and addressing human factors problems in job and workplace design, the workplace environment, and training.

Apparel manufacturing is a labor-intensive process requiring 20 or more distinct steps for the completion of a garment. Each of these steps involves the operation of one or more kinds of power equipment such as specialized cutting and sewing machines. The steps require significant amounts of manual manipulation as fabric bundles, individual pieces, and unfinished garments are transported, inserted into, and guided through the machines. These operations are highly repetitive and visually demanding. Many of the jobs involve complex psychomotor skills that require three to four months of training and practice to achieve acceptable performance levels. Wages are typically on a piecerate basis. While the introduction of high-technology, semi-automated equipment to the manufacturing floor can help to ameliorate some of the problems, it will create new difficulties while changing the basic nature of most jobs very little.

The purpose of this program was to explore human factors problems in typical plants manufacturing military uniform trousers or similar trousers and to recommend cost-effective interventions for the identified problems. Subsequent work was planned to test recommended solutions in conventional manufacturing environments and to project and address human factors issues in the advanced technology uniform manufacturing plant of the future.

Inadequate job design and badly designed workplaces can create musculoskeletal discomfort or disorders in assembly workers due to awkward working postures, excessive reaches, excessive manual manipulation, excessive strength or endurance requirements, or combinations of these factors. Repetitive work, for example, may result in health problems such as cumulative trauma disorders that increase absenteeism and reduce productivity. One such disorder, carpal tunnel syndrome, is associated with repetitive forceful movements of the hand with deviated wrist postures and is characterized by pain and numbness in the hand and wrist.

1 This Research is sponsored by the Department of Defense, Defense Logistics Agency. Mr. Dan Gearing serves as the program Technical Monitor. The authors would like to thank Ms. Lisa C. King, Ms. Catherine Braza, and Mr. David Norman for their assistance.

Data about such musculoskeletal discomfort can provide important clues about workplace and job design problems. In addition, this information is useful in establishing priorities for automation of particular manual operations. If operations that are particularly stressful to the musculoskeletal system can be modified to eliminate discomfort or can be successfully automated, it will be easier to retain experienced, skilled workers and to maintain their productivity.

Careful design of tools and equipment for the workplace requires the engineer to obtain and apply data on the relevant physical dimensions of the worker population. Such data, however, are not readily available for most civilian populations including apparel cutting and sewing operators. Casey (1989) demonstrated the imprecision in equipment design that can result from attempts to apply the few existing civilian anthropometric data bases to the design of equipment for specific user populations. Data were needed on the physical dimensions of the apparel manufacturing operator population.

Training of manufacturing workers is a significant issue because of the high rate of turnover experienced by the industry and because of the lengthy learning time required to reach proficiency on many of the more demanding jobs. Presumably, increasing the effectiveness of training could provide multiple benefits by enabling workers to achieve more productivity and, consequently, more rewarding rates of pay earlier in their careers. It might also help to reduce the incidence of hazardous work habits that contribute to the frequent musculoskeletal pains and disorders found in the industry.

Method

Site visits were conducted at three plants that manufacture trousers in the Southeastern United States. The plants were selected to provide a range of company sizes (50 - 600 employees) and degrees of automation that is typical of the industry. Two of the three produce only military uniform trousers while the third produces a range of casual and dress trousers.

A total of 132 operators at the three sites volunteered to take part in the research. Of these operators, 123 were female and 9 were male. The participants had a mean age of 40 years and a mean of 8.5 years experience in the apparel industry. These demographics appeared to be representative of the operator populations of the apparel manufacturing plants we visited. Data collection took place before and after work and during lunch breaks.

Anthropometric measures were taken on all participants. Subjects were measured on each of ten dimensions related to workstation and operator interface including (1) stature, (2) eye height, (3) shoulder height, (4) standing elbow height, (5) arm length, (6) seated height, (7) popliteal height, (8) thigh thickness, (9) hand length, and (10) hand breadth. The measuring devices were a GPM Model 101 anthropometer and a GPM Model 106 spreading caliper. Subjects were measured with shoes removed and were wearing lightweight summer clothing.

Confidential interviews were conducted with all subjects using a structured interview form. The interviews covered (1) demographic information, (2) musculoskeletal injuries or discomfort, (3) the job environment, (4) characteristics of the workstation, chair, and job, (5) training, (6) work schedule, (7) other factors that might influence safety or productivity. To promote uniformity, most of the interview consisted of

YES/NO, multiple choice, or short answer questions. The final two questions, however, were open-ended and interviewers attempted to draw out more detailed information on previous responses or suggestions for job improvement.

Based on indicators of human factors problems such as high turnover or absenteeism, a history of complaints or injuries, or unusually high skill requirements, 14 target jobs were selected for more detailed study. Video registration and analysis using methods described by Keyserling (1986) and Melin (1987) were used to record and score hand, wrist, back, and neck position and activity of the selected operators during sewing operations. Measurements of the "problem" workstations were also taken.

Ambient noise and noise levels at the operators' ears were measured at a sample of workstations in each facility using a General Radio Model 1982 Sound Level Meter. Measures of ambient illumination and the illumination at the workstation point-of-operation (POO) with supplementary task luminaires (if normally used) were taken with a Gossen-Panlux light meter.

Results

Anthropometry

Table 1 summarizes the mean, 5th percentile, and 95th percentile values for each of the ten measures taken of the 123 female volunteers in the study. Compared to other sources of anthropometric data on civilian females (Webb Associates, 1978; Society of Automotive Engineers, 1983) the population of sewing operators appeared to be slightly smaller (based on stature, seated height, etc.) and slightly heavier (based on thigh thickness).

TABLE 1: ANTHROPOMETRIC DATA FOR 123 FEMALE SEWING OPERATORS

DIMENSION	MEAN	S.D.	5TH %ILE	95TH %ILE
STATURE (cm)	160.7	6.4	150.2	171.2
EYE HEIGHT (cm)	150.6	6.2	140.5	160.7
SHOULDER HEIGHT (cm)	133.3	5.9	123.7	143.0
ELBOW HEIGHT (cm)	103.1	4.5	95.6	110.5
ARM LENGTH (cm)	69.8	3.6	63.9	75.6
SEATED HEIGHT (cm)	84.2	3.5	78.5	90.0
THIGH THICKNESS (cm)	15.0	2.1	11.6	18.5
POPLITEAL HEIGHT (cm)	41.7	2.5	37.6	45.8
HAND LENGTH (cm)	17.5	0.9	16.1	18.9
HAND BREADTH (cm)	7.8	0.4	7.1	8.5

Postural Discomfort

As shown in Figure 1, about half of all interviewed workers reported that they at least "sometimes" suffer from pain in their upper back (52%), neck (49%), and right hand (48%). Based on interviews with methods engineers and observation, the jobs were classified as low, medium, or high in manual manipulation. Operators in high manipulation jobs reported a substantially higher frequency of pain in the hands, arms, shoulders, middle and upper back, and neck than did the operators in the low manipulation jobs. These operators also reported a high incidence of pain and numbness in the hands and fingers, one common early symptom of carpal tunnel syndrome.

Operators in standing jobs, as might be expected, reported significantly greater amounts of foot discomfort and less discomfort in the back and neck than did operators in sitting jobs. An additional trend found less discomfort being reported by older workers than younger workers. This might be attributed to a reticence on the part of older workers to report job-related discomfort, to a reassignment of older workers to less demanding jobs, or to a self-selection of individuals susceptible to cumulative trauma disorders out of the apparel manufacturing workforce.

Video analyses were performed on a sample of 30 operators of varying experience levels. Differences between experienced and novice operators were examined. Experienced operators were found to make fewer and more ballistic hand movements during a given cycle. Because experienced operators performed many more cycles per unit of time than the novice workers, however, their level of hand activity was greater.

Video analyses were performed on the postures of these sewing operators. During the machine cycle, 40 percent of these operators stooped forward (i.e., torso flexion) at least 20 degrees during the cycle. Sixty percent of the operators tilted their heads forward more than 20 degrees throughout most of the cycle. Such hunched postures have been cited as a factor associated with fatigue and discomfort (Grandjean, 1988).

The prevalence of this hunched-over working posture is probably the result of two factors, visual requirements and workstation geometry. The quality requirements of the sewing job require clear visual perception of the POO. Yet, measures indicated that average illumination levels at the POO (168 footcandles) were substantially lower than the recommended values (300 footcandles). During interviews, 36 percent of the operators complained of inadequate illumination.

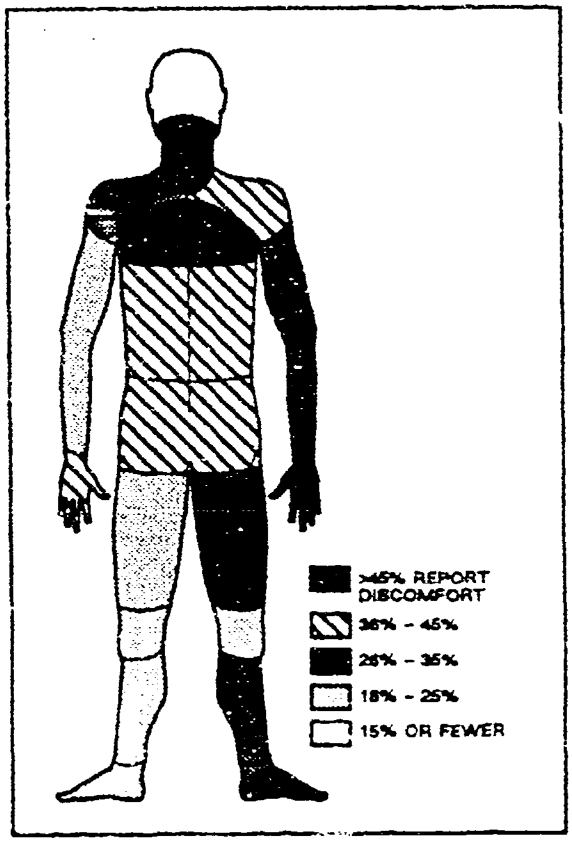


Figure 1. Percentage of Workers Reporting Musculoskeletal Discomfort by Body Area

Measures of the typical workstation indicated that the mean position of the foot treadle was only 15 cm from the proximal edge of the work surface. Commonly, operators positioned the chair back away from the work surface in order to allow an optimum knee angle of 110 degrees or greater. Because of the resulting distance from the chair to the work surface, the average reach required was 84 cm, nearly double the recommended 45 cm (Eastman Kodak, 1983). From this chair position, operators were required to lean forward 25 degrees or more in order to have adequate visual and manual access to the POO.

Training

On-the-job training was the standard at all three sites. Only one plant used a dedicated training staff; at the other two, training was provided by the floor supervisors. None of the trainers at any of the sites had any formal instruction in effective training methods. The quality of instruction and performance feedback varied between and within plants. Only 17 percent of the operators reported that they had ever been provided with any kind of written or graphical training or job performance aid. Thirteen percent reported that some use of videotape was made in their training, most often during cross-training for a second position.

Discussion

Among sewing operators, complaints of musculoskeletal discomfort centered in the torso and upper extremities. Analyses suggested that these complaints can be partially addressed by adequate workstation and chair adjustability as well as by improved illumination.

Research is needed to determine the actual prevalence of chronic trauma disorders in trouser manufacturing. Our results suggest that the prevalence may be considerably higher than reported. The operations studied here were all hand-intensive. Reducing the frequency of high risk hand motions will require modification of workstations and methods; eliminating the use of these motions altogether will require automation.

More effective training could improve productivity and increase safety. We would recommend improvements in four areas: (1) improved training for the training staff, (2) increased use of training aids, (3) better performance feedback during training, and (4) more emphasis on injury prevention and posture training.

References

Casey, S. M. (1989). Anthropometry of farm equipment operators. Human Factors Society Bulletin, 32(7), 1-3.

Eastman Kodak Company. (1983). Ergonomic design for people at work, Vol.1. Belmont, Ca.: Lifetime Learning Publications.

Grandjean, E. (1982). Fitting the task to the man: An ergonomic approach. London: Taylor and Francis, Ltd.

Keyserling, W. M. (1986). A computer-aided system to evaluate postural stress in the workplace. American Industrial Hygiene Association Journal, 47(10), 641-649.

Melin, E. (1987). Neck-shoulder loading characteristics and work technique. Ergonomics, 30(2), 281-285.

Society of Automotive Engineers. (1983). USA human physical dimensions. SAE J833 DEC83. Warrendale, Penn.: Author.

Webb Associates. (1978). Anthropometric source book, Volume I: Anthropometry for designers. Washington, DC: National Aeronautics and Space Administration. (NTIS No. N79-11734).

Macro-Ergonomics and Exchange Processes

John Landry
Iowa State University
Ames, Iowa 50011

Abstract

Hendrick (1988) introduced "macro-ergonomics" to direct attention toward human-human interaction in addition to human-machine interactions. This conceptual reorientation is advocated as a foundation for improving organizational performance. This paper offers some conjecture about the relationship among learning, decision making, communication, and organizational performance in the context provided by macro-ergonomics.

Macro-ergonomics refocuses attention toward the transactions occurring between organization members. Transactions between entities range from exchanges of labor and money to exchanges of power and influence (Jones, 1983). Moreover, Levine and White (1961) claim that exchange is a useful framework for studying organizations. This paper examines macro-ergonomic exchanges from three perspectives in order to develop conjectures about learning, decision making processes, communication processes, and organizational performance.

Three Perspectives on Exchange

The first perspective looks at organizations from a cognitive perspective. This view assumes that mental maps form the basis of organization reality and roles (Argyris and Schon, 1978). Mental maps, either public or private, explicit or implicit, are the shared beliefs of organizational members used to guide actions and attach meaning to events. Learning occurs as the content and linkage between concepts in these maps changes (Argyris and Schon, 1978).

Interacting with the environment allows members of an organization to create new interpretations of their circumstances (Argyris and Schon, 1978). New meanings are stored as mental maps that are transferred to the organization through communications systems and guide the members' behaviors. For example, SOPs, policy manuals, procedures, contracts, and laws describe how organizational roles can or must respond.

Organization theory generally accepts that an organization's primary objective is to insure organizational survival by seeking a "fit" with the environment (Thompson, 1967; Tosi, 1984). These exchanges with the environment, which are required for learning, are the primary source of uncertainty managers must accommodate (Thompson, 1976). High performance and strategic advantage are achieved by fitting the organization to the environment and minimizing the costs of the exchanges required to achieve a fit (Thompson, 1967; Tosi, 1984).

The uncertainty created by environmental interactions of open systems prompts rational organizations to forecast problems and develop responses to anticipated problems. Unfortunately, many problems cannot be determined a priori because of incomplete or unknown cause-effect relationships. Leaders and managers are called upon to work out some mechanism to handle problems arising from unexpected situations. Kolb (1974) observed that learning and problem solving are different perspectives of the same issue.

A second view of the macro-ergonomic model develops from the notion that leaders are called upon to solve problems by making decisions. Simon (1976) observed every decision has two components, facts and values. Facts are empirically testable but values are derived from morals or personal preferences. Another perspective on this classification is provided by Laughlin's decision making research (McGrath, 1984). Decision making tasks may be classified as either intellective tasks or judgmental tasks. Intellective tasks are those which have a demonstrably "correct" answer; judgmental tasks do not. Intellective task "correctness" is established as a social phenomena. For example, accepting the rules of evidence established by a science as an indicator of truth or following legitimate procedures within mathematics, engineering, and logic are considered indicators of correctness. In the judgmental category, a "correct" answer is based upon consensus among peers, preferred outcomes, moral judgments, and values (McGrath, 1984).

A third perspective on exchanges is that communication is the transfer of information between entities (Daft & Huber, 1987; Weick & Browning, 1988). The transfer can be viewed from two different communication paradigms, argument and narration (Weick and Browning, 1988) . Argument requires rational evaluation based upon a coherent set of rules established a priori. Technical expertise, public debate, common goals, and demonstration of truths are valued in this form of communication. On the other hand, narrative communication assumes "values are ultimately persuasive ... soundness is the essential standard of reasoning. (Weick & Browning, 1986: 249)"

Since the objective of an organization is to achieve a fit with its environment, making a correct decision relative to the problem addressed is essential. Intellective and judgmental tasks can be reframed in terms of their source of validation being either internal or external to the exchange. Therefore, judgmental, or value based decision making, has significant organizational implications since it establishes the idiosyncratic organizational understandings which are central to an organization's uniqueness. Effective judgmental decision making supports developing a strategic advantage by attaching a unique meaning to events and sharing that meaning through culture, shared mental maps, and interpretive schema.

Achieving strategic advantage or maintaining superior performance is a ubiquitous goal of organizations. The macro-ergonomic framework focuses on the human interaction to achieve that goal. The three notions discussed add several additional insights into exchange processes.

1. Exchange is a fundamental process of organized activity in an open system.
2. Exchange is necessary for learning but it creates uncertainty.
3. Uncertainty is reduced by increasing information; information is

created when data are applied to influence a choice, that is, make a decision.

4. Decisions rely on one of two standards for judging performance, rules or social consensus.

The following is a conjecture about how the insights may be applied to macro-ergonomic processes and extended to subsume micro-ergonomic processes.

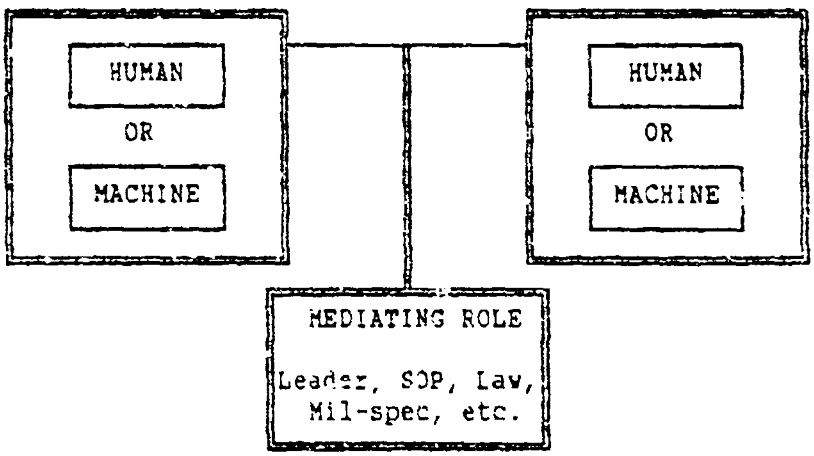
Linking Macro-Ergonomics and Exchange

A link between the two models can be made by assuming that all exchanges must be mediated by a third party (Weick, 1969). At a pragmatic level this assumption can be tested by asking, "What is the role of a leader or supervisor?" Their roles assign them responsibility over certain domains. They may either establish procedures or make decisions when procedures are not available. If the problem is outside of their domain, they can refer it to others and the cycle repeats. They act to mediate exchanges within their domain.

The micro-ergonomic model can be tested in a similar manner by asking, "What is the role of standards or procedures?" They guide the interaction. The human factors approach is used to establish procedures and machine standards. In fact, the same rationale could be extended to standards imposed on one machine interfacing with another. These concepts are represented in Figure 1.

Figure 1.

Macro-Ergonomic Exchange Model



This model suggests several facets that a manager must consider as either a designer of a macro-ergonic exchange or as a particpant in the exchange. First, one must consider control of the exchange within the assigned area of responsibility. If the manager's role is to design the exchange, then he or she must decide if they can, in fact, design the control process. Control decisions should be reliquished to others when they are outside the manager's area of responsibility or when the rules have been established by external governing bodies.

Second, one must consider the processes used establish the exchange performance. In circumstances where the performance can be judged according to preexisting standards, those standards may be used. Determing the availability of standards may not be easy since it may require using experts or finding difficult to comprehend procedures. On the other hand, the alternative is to "reinvent-the-wheel" which is an expensive and time-consuming process.

Third, one must consider the nature of the communication process. In general, organizational theorists espouse the benefits of using a rational approach to communications. However, the problem may not have the factual basis required to use this communication strategy. In this circumstance, the strategy is inappropriate for the problem. In other words, the decision maker may fail to "fit" the organizational process to the problem environment, which may lead to lower performance.

The three components of the decision process associated with a macro-ergonmic perspective are summarized in Table 1.

Table 1.

Macro-Erogomic Decision Processes

Domain of Decision Control	Internal	External
Evaluation and Decision Making	Judgmental	Intellective
Communication Process	Narration	Argument

As shown in Table 1., the essential feature of the exchange is origin of the the control process.

Summary

Organizations exist as sets of exchanges guided by their mental maps and maintained by high performance decision making. Exchange provides a useful model for integrating learning, decision making, communication, and performance evaluation.

References

Argyris, C & D. A. Schon (1978). Organizational Learning: A Theory of Action Perspective, Addison-Wesley, Reading, MA.

Daft, R. L. & G. P. Huber (1987). "How Organizations Learn: A Communications

Framework," Research in the Sociology of Organizations, 5, 1-36.

Hendrick, H. (1988). "Macroergonomics: A System Design Approach for Developing More Effective Military Systems," Proceedings Psychology in the Department of Defense, Eleventh Symposium, April 13-15, 57-61.

Jones, G. R. (1983). Transaction Costs, Property Rights, and Organizational Culture: An Exchange Perspective. Administrative Science Quarterly, 28, 454-467.

Kolb, D. A. (1974). "On Management and the Learning Process," in Kolb, D. A., Rubin, I. M., and MacIntyre (Eds), Organizational Psychology, Prentice-Hall, Englewood Cliffs, NJ.

Levine, S. & P. E. White (1961). Exchange as a Conceptual Framework for the Study of Interorganizational Relationships, Administrative Science Quarterly, 5, 583-601.

McGrath, J. E. (1984). Groups, Interaction and Performance, Prentice Hall, Englewood Cliffs, NJ.

Simon, H. A. (1976). Administrative Behavior, 3rd Ed., The Free Press, New York, NY.

Thompson, J. D. (1967). Organizations in Action, McGraw-Hill, New York, N.Y.

Tosi, H. L. (1984). Theories of Organization, John Wiley & Sons, New York, NY.

Weick, K. E. (1969). The Social Psychology of Organizing, Addison-Wesley, Reading, MA.

Weick, K. E. and Browning (1986). "Argument and Narration in Organizational Communication," Journal of Management, 12(2), 243-259.

Human Responses to Weightlessness

Orwyn Sampson, PhD
United States Air Force Academy

Abstract

An introduction to the phenomenon of weightlessness is presented in terms of its cause and its effects on the human body. Following a discussion of just what weightlessness is, its major physiological effects are reviewed. These include: 1) altered body orientation stemming from the removal of a major source of sensory stimulation of the vestibular apparatus contained within the inner ear; 2) body fluid shifts resulting in compensatory cardiovascular system responses; and 3) changes in body composition, especially muscle and bone, with prolonged exposure to weightlessness. Each of these effects is discussed with the goal of understanding its etiology. In addition, implications for human movement in space and on earth are given.

WEIGHTLESSNESS

There is a common misunderstanding about weightlessness. We have developed the notion that weightlessness and space go together, that people become weightless whenever they go beyond the earth's atmosphere.

We have acquired this notion naturally. For years now, we have become accustomed to seeing astronauts and cosmonauts "floating around" inside their spacecraft once they achieve earth orbit. And, since the orbit altitude, usually 150 - 300 miles, is beyond our atmosphere, it is natural to associate weightlessness with a lack of atmosphere; i.e., with space. This association is natural, but in fact, it is only coincidental.

According to Newton, every object in the universe attracts every other object. The amount of attraction that exists between two objects is directly related to their masses and inversely related to the distance they are from each other. Because gravity is universal, a state of absolute weightlessness could exist only if you were present in an empty universe. And, since that is hardly feasible, our conception of weightlessness needs to be adjusted. To understand it properly, we need a frame of reference. To illustrate: on earth, we experience our mass as weight because the earth pushes up on us opposing gravity's pull. If the earth suddenly gave way, we would no longer experience our mass as weight because there would be nothing to oppose the pull of gravity.

When space travelers appear to be weightless in orbit, as we have just been reminded, they are <u>not</u> free of gravity. Instead, they are falling. They are still being attracted to the earth and are falling toward it, but because they have a forward velocity (of some 17,000 mph), they keep missing it. And, since their spacecraft is also falling, using <u>it</u> as a frame of reference, they are indeed weightless. In other works, they are weightless <u>relative to their spacecraft.</u>

Therefore, what we have come to regard as weightlessness is simply a state of free fall in which our frame of reference is moving in the same direction and at the same speed we are. This being the case, it is now easier to see that weightlessness has nothing to do with space per SE. Indeed, we experience weightlessness within our atmosphere every time we are free from support and are allowed to fall. When a pilot noses his plane over after a gradual climb, or an elevator descends abruptly, we experience weightlessness. The only difference is in the duration. Because this is such a foundational concept, let us illustrate it another way be considering an implausible situation.

Suppose we built a skyscraper 250 miles high. The person living in the penthouse, to be sure, would require supplemental air to breathe because our atmosphere does not support life much beyond about five miles elevation. But. mark this, he or she would not be weightless! In fact, their weight would be diminished by a mere 12%, and that not because of space, but because they are now 250 miles further from the center of the earth.

As we move out from the earth, our weight indeed diminishes. But, distance from the earth is not the only factor: The penthouse dweller and the space traveler are the same distance from earth, but one still has weight, the other does not. The difference, of course, is accounted for by the fact that the space traveler and his spacecraft are both falling.

To put this, hopefully, in even clearer perspective, here is a view of the earth from the moon. Its diameter is 8,000 miles. The shuttle altitude is 250 miles. And, on an 8,000-mile scale, 250 miles is not very far out into space. Clearly, these space travelers have not escaped the earth's pull of gravity!

What, then, is the correct relationship between space and weight? The short answer is that space facilitates free fall better than air does, so it is easier to orbit in space than within our atmosphere. And, that is the key. While it is possible to orbit within our atmosphere, air provides so much resistance that the cost of achieving and maintaining a forward velocity of 17,000 mph would be astronomical (and I know it's a pun). Conversely, in space, where there is practically no impediment to movement, speeds of that magnitude or greater, once achieved, can be maintained with no additional power.

With the advent of space travel, mankind faces a radically new experience, a longterm gravity-diminished existence.

As of the first of this year, more than 200 people have traveled beyond our atmosphere with a collective experience of 21.4 manyears in space.

Eventually, with the advent of space station and trips within and beyond our solar system, the weightless experience will be extended perhaps indefinitely.

Weightlessness is an exhilarating experience, but there are costs associated with it.

PHYSIOLOGICAL EFFECTS

Exposure to microgravity has three major effects on human bodies. These include:

1. Changes in body-environment orientation (nervous system effects).
2. Changes in body fluid distribution (circulatory system effects).
3. Changes in body structure (primarily effects on muscle and bone).

Nervous System

A primary function of the nervous system is basic orientation--orientation of your body with respect to your environment. Orientation involves both sensory and response activities on your part. The primary sensory mechanisms include: vision, the inner ear, and somatosensory receptors (including the kinesthetic sensors). The response activities include: eye movement, changes in head position, and altered posture.

When the environment varies beyond certain limits, the sensations can be disorienting and may produce motion sickness. This occurs in a weightless environment.

More than half of all space travelers get sick in microgravity. The condition is labeled space motion sickness (SMS). Some of the symptoms include nausea, headache, anorexia, and vomiting. SMS starts almost immediately for some astronauts and can last up to four days.

A study of generic motion sickness reveals some interesting facts:

1. It is not limited to man. In fact, nearly every species exhibits susceptibility including horses, cows, monkeys, chimpanzees, seals, various kinds of birds, sheep, and cats. Even fish, under the right conditions, exhibit motion sickness.
2. The only common characteristic of all the motions producing the syndrome is varying acceleration to the head and one fact stands out: Individuals not possessing a functioning vestibular labyrinth (e.g., some deaf mutes) are not susceptible to motion sickness.

The vestibular labyrinth, or vestibular apparatus, is located in the inner ear and is continuous with the structure responsible for hearing, the membranous cochlea. It contains two types of senory receptors: otolith organs which sense linear accelerations (including gravity); and, semicircular canals which sense angular accelerations.

The otolith organs (viz, the utricle and saccule) are linear accelerometers and provide the central nervous system with a continuous estimate of head position with respect to the upright. The semicircular canals are rotary accelerometers and detect rotations in the three basic planes or reference. Microgravity alters the activity of the otolith organs but not the semicircular canals. This leads to senory conflict and space motion sickness.

Although the exact nature of SMS remains a mystery, it is not a permanent condition and is treatable with drugs and biofeedback techniques.

Circulatory System

Microgravity has two effects. One is an immediate effect of redistributed body fluids resulting from decreased hydrostatic pressure. This leads to thin legs and a puffy face. This headward shift of fluids is interpreted by the body as an increase in blood volume. The circulatory system adapts to this sensed "volume overload" quickly. Within the first day or two, up to two liters of excess body fluids are excreted and the number of red blood cells is also reduced.

The second effect of microgravity is a longterm effect in which the heart muscle atrophies and certain changes are seen in the blood which are described collectively as "spaceflight anemia". Because of these changes, there is increased likelihood of shock if hemorrhage occurs. And, upon reentry to a 1-G environment, orthostatic tolerance is significantly diminished. Recovery may take days to weeks.

Measures employed to counter the adverse effects of microgravity on the circulatory system have taken a variety of forms and include:

1. Inflight aerobic and strength training exercises
2. Lower body negative pressure
3. Oral rehydration (within two hours of reentry)
4. Drugs
5. Hormones
6. And the use of anti-G suits postflight

Muscular Systems

Shortterm space flights do not impact muscles to a significant degree--there is a decrease in body weight but this is mostly water. In longterm flights, both body water loss and negative nitrogen balance are seen.

The loss in muscle occurs mainly in the lower extremities. In addition to the loss of muscle mass, the muscles tend to "speed up". In other words, the atrophy appears to occur primarily in slow-twitch fibers.

Muscle atrophy can be effectively countered by means of exercise, electrical stimulation, and hormones.

Skeletal System

Osteopenia (literally, "poor bones") is a consistently reported physiological response to spaceflight. It occurs because the ongoing processes of bone resorption exceed those of bone formation.

Spaceflight osteopenia is more severe than bedrest osteopenia and may not be completely reversible. Bedrest osteopenia plateaus, but spaceflight osteopenia has one element that does not. Spaceflight osteopenia is even more serious than post-menopausal osteopenia. A related factor in osteopenia is muscle atrophy because of the stress effect muscles have on bones. Postflight there is increased likelihood of bone fracture and an increased likelihood of kidney stone formation.

Two countermeasures have shown some benefit in retarding bone loss. These include exercise and diet, but osteopenia is not totally prevented or reversed using either countermeasure and it remains a serious effect of microgravity.

IMPLICATIONS

What does all this mean for us? I think there are three major implications:

1. One has to do with moving oneself in a microgravity environment. We use free body diagrams all the time to describe force and torque effects. In microgravity, we become "free bodies.
2. A second implication has to do with moving other objects in space. this calls for a concentrated review of, and experience with, the physics of movement in terms of forces, centers of gravity, torques, mass, etc.
3. And a third implication has to do with training-programs; programs and research for countering the adverse effects of weightlessness and preparing space travelers for skilled, efficient movement in microgravity. There is great need for creative, innovative research in this area.

Gender Identity Among Air Force Female Aviators*

Major Karen O. Dunivin, Ph.D.
USAF Academy

Abstract

This article examines one facet of socialization of women in the male-dominant world of flying by studying Air Force female aviators' gender identity. According to the Bem Sex Role Inventory, female aviators display more masculine and androgynous identities than other women surveyed in both military and civilian studies. Furthermore, they are more likely to shift toward masculine identities. Interestingly, female aviators in Tactical Air Command (TAC), the predominantly male fighter command, display the greatest proportion of masculine identities. These findings suggest that female aviators adopt more masculine identities and socially-desirable masculine characteristics in order to succeed in the male-dominated world of flying.

Most gender studies of the workplace explain observed differences between the sexes (e.g., behaviors or attitudes) by noting individual differences between men and women (e.g., physiological or psychological differences), or citing contextual factors (e.g., organizational structure or socialization) as causes for the observed differences. Additionally, studies of military members have examined gender differences and role socialization -- does the masculine world of the military provide similar socialization experiences for men and women (Moskos, 1985; Devilbiss, 1985)? Does the military reflect a gender-pluralistic culture, or is there a dominant masculine culture that pressures female members to adopt more masculine attitudes, orientations, and identities? Gender is an important element in shaping a person's self-image and social identity (Tischler et al., 1986). Gender identity is culturally defined (i.e., "appropriate" male and female behaviors vary by cultures) and is learned through role socialization. It is an acquired identity that changes over individuals' life cycles and cultural contexts.

This article examines gender identity among Air Force female aviators -- those who work in the predominantly masculine world of flying. Analysis of gender identity of women in masculine-dominant cultures may provide insight into the relationship between organizational socialization and individual adaptation. The military has been described as a distinct culture that promotes masculine identities, attitudes and behaviors (Segal and Segal, 1983). Furthermore, aviation has historically been considered men's work (Wood, 1982). However, military women are born female (ascribed status) and develop appropriate gender identities. Therefore, conflict between organizational and gender socialization may affect gender identity. Thus, women may sacrifice their conventional gender identity (i.e., femininity) for a masculine identity that is reinforced and rewarded in the masculine world of

* Views expressed in this paper are those of the author and may not reflect those of the U.S. Air Force or the Air Force Academy.

the military. This pattern may be more pronounced for women who perform sex-prescribed male work such as flying.

A common measure of gender identity is the Bem Sex Role Inventory (BSRI). Respondents characterize themselves on culturally desirable traits which classify their sex-role (or gender) identity into one of four categories: masculine, feminine, androgynous, and undifferentiated (Bem, 1974). Those categorized as masculine see themselves as self-reliant, independent, and athletic (culturally-defined masculine traits); those classified as feminine describe themselves as compassionate, loyal, and understanding. Androgynous individuals describe themselves as having both strong masculine and feminine identities. Undifferentiated persons are low both on masculine and feminine scores -- there is no distinguishable pattern of gender identity.

Androgyny is viewed as the most desirable profile because it depicts persons who attribute both masculine and feminine traits to themselves (Bem, 1974; Cook, 1985). In short, androgynous persons have behavioral flexibility to positively react to different situations. They can use masculine <u>or</u> feminine characteristics (e.g., be forceful or compassionate) as the situation warrants -- they do not feel constrained by rigid sex-role norms. Conversely, undifferentiated persons have no well-defined gender identity. Indeed, some research suggests that persons categorized androgynous and masculine tend to be high on self-esteem compared to feminine or undifferentiated persons (Bem, 1977). Spence and Helmreich (1978) describe a continuum from highest to lowest self-esteem to be consistently associated with the four groups in this order: androgynous, masculine, feminine, and undifferentiated.

Method

Data reported here are part of a larger study of female Air Force aviators who are members of Women Military Aviators, Inc. (WMA). The respondents completed a 140-item questionnaire that included various demographic and attitudinal questions. This study's sample (53 percent return rate) includes 120 active-duty Air Force female aviators who represent approximately 22 percent of the total population of Air Force female aviators. Ideally, all Air Force female aviators should be surveyed; however, time and cost constraints prohibited such a comprehensive study. Some may argue that the sample may reflect unique attitudes of those who belong to WMA. However, in my interaction with female aviators (both WMA members and nonmembers), I have found no unique profile specific to either group. Therefore, the sample is assumed to represent Air Force female aviators in general. And demographically, the respondents are diverse in terms of age, years of service, military rank, and major command (MAJCOM) affiliation.

Using the BSRI, two scores are calculated: masculinity and femininity. Research consistently reports a median for both scales at approximately 4.9 (Hyde and Rosenberg, 1980). As discussed earlier, individuals are classified as masculine, feminine, androgynous, or undifferentiated. For example, if one's masculinity and femininity scores (means) are above 4.9 (approximate medians), the respondent scores "high" on both scales and is classified androgynous; one who scores "low" on both scales is termed undifferentiated. Previous research reveals that the BSRI is a valid, reliable measure of gender identity (Cook, 1985).

Results

Table 1 presents comparative studies reporting BSRI scores (this study's data are listed on the top row). The overall mean scores of masculinity and femininity for the surveyed Air Force female aviators are 5.5 and 4.9, respectively. Using the accepted normative median scores of 4.9 for both scales, Air Force female aviators do not have conventional gender identities (i.e., feminine identities). Instead, they have more masculine and androgynous identities (42 and 45 percent, respectively). Conversely, few female aviators have feminine identities (10 percent) which sharply contrasts with studies of nonmilitary women (34 to 52 percent).

TABLE 1

PERCENT DISTRIBUTION OF WOMEN IN SEX-ROLE CATEGORIES

	N =	Undifferentiated	Feminine	Masculine	Androgynous
BSRI[a]	120	3%	10%	42%	45%
BSRI[b]	35	6%	8%	26%	60%
BSRI[c]	290	20%	34%	16%	29%
BSRI[d]	369	19%	47%	10%	24%
BSRI[e]	165	13%	52%	15%	20%
BSRI[f]	85	15%	49%	14%	21%

Sources:

a Dunivin, 1990 (Air Force female aviators)
b Dunivin, 1988 (Air Force female officers)
c Bem, 1977 (male and female college students)
d Hoffman and Fidell, 1979 (adult women, age 20-59)
e Hyde and Phillis, 1979 (males and females, age 13-85)
f DeFleur and Warner, 1987 (Air Force Academy female cadets)

Notes:

1. Percentages are computed by rows.
2. Percentages may not add to 100% because of rounding.

Most striking are the differences in percentages for the masculine and androgynous categories. Forty-two percent of female aviators surveyed have masculine identities, compared to the other studies (except Dunivin, 1988) which range from 10 to 16 percent. Moreover, 45 percent of the women have androgynous identities, again, higher than the nonmilitary samples. Cumulatively, 87 percent of the surveyed women have masculine or androgynous identities, compared to the other samples (except Dunivin, 1988) where the two identities only account for 34 to 45 percent of the sample -- a startling difference.

Comparing two similar populations (i.e., Air Force female officers [Dunivin, 1988] and these aviators), there are similarities and differences. The median scores are similar (Dunivin [1988] reported medians of 5.6 masculinity; 5.05 femininity). Likewise, the percentages are similar for the combined categories of masculine and androgynous identities. However, female aviators have more masculine identities (42 percent) than Air Force female officers (26 percent). This difference suggests a shift toward more masculine identities for female aviators than female officers in general.

Examining gender identity among the various subgroups, there are few discernible patterns. Factors such as military rank, age, commission source, marital status, and career intentions show no distinct patterns. However, one pattern appears. Female aviators in MAC and SAC have a balance between masculine identities (38 percent and 30 percent, respectively) and androgynous identities (52 percent and 54 percent, respectively). However, 80 percent of the surveyed female aviators in TAC have masculine identities and none have feminine or undifferentiated identities. This pattern may suggest that military women in the most masculine culture of TAC's fighter world adopt almost exclusively masculine identities.

Discussion

When examining BSRI scores in the different studies (Table 1), some differences may be attributed to the diverse populations of the studies. For example, two studies use college students as subjects, a rather specialized segment of a society's population. Still, the findings suggest some important differences with alternate plausible explanations. First, none of the studies (except Dunivin, 1988) measure professional career women's gender identity. For the most part, career women in management positions (including Air Force female aviators) work in an organizational culture in which leadership, assertiveness, and other male-attributed qualities are viewed as positive traits. It is plausible that women in such work environments adopt more masculine identities and socially-desirable masculine characteristics in order to succeed as leaders or aviators. However, some women have strong feminine identities (developed over years of gender role socialization); therefore, they adopt more androgynous identities.

A second point is that this study includes only female aviators who remain in the Air Force. The rigors of a military lifestyle may have driven out women with lower self-esteem (as associated with feminine and undifferentiated gender identities). Furthermore, the military environment promotes masculine behaviors and attitudes, conducive to women with masculine or androgynous identities versus feminine or undifferentiated identities. It may be argued as well that some women with high self-esteem (as associated with masculine and androgynous gender identities) may have separated from the Air Force after they experienced sexism, career disappointments, or family conflicts. However, we can only survey the female aviators who remain in the Air Force.

A third explanation may be that the military and aviation attracts women with unconventional gender identities. Women who choose flying as an occupation may be more masculine in their orientations and interests. However, without longitudinal data (i.e., a measure of women's gender identity before their entry and during their stay in the Air Force), there is no way to determine if women have more masculine or androgynous identities prior to military socialization.

A fourth interpretation may be that female aviators initally have stronger feminine identities, but alter their gender identities to conform to masculine culture, especially in the flying world. If gender-role socialization is a lifelong process, gender identity can change. Two previous studies may support this assertion. DeFleur and Warner (1987) found that 49 percent of USAFA female cadets had feminine identities. Some of these women became officers and aviators and may be participants in this study. In comparing this study with DeFleur and Warner, there appears to be a shift from feminine identities toward masculine and androgynous identities for female officers, and more masculine identities for female aviators. Furthermore, in comparing Dunivin (1988) and this study, there appears to be a transition from androgynous identities of female officers to masculine identities for female aviators. It is plausible that as female aviators adjust to the flying world they adopt a more masculine identity which is fostered among aviators. Again, only longitudinal data can determine whether women change their gender during their military careers.

Conclusions

In comparison to women in general, Air Force female aviators have nontraditional gender identities. However, the flying world promotes masculinity, and the women may acquire a masculine identity or blend their gender identity to adopt an androgynous identity in order to succeed. Consequently, the women place less importance on their feminine identity in order to survive and thrive in a masculine culture.

These patterns make sense if we recognize that workers (male or female) are socialized to accept the organization's values and standards. If the military primarily promotes masculine values, behaviors and attitudes, female members must adopt such values and standards in order to work and build successful careers.

However, there remains the question whether women should adjust to the Air Force's male world, particularly in aviation. Or should the culture reflect its pluralistic composition? Some contend that the culture should reflect the organization's purposes, not its racial, sex, or social composition. Others argue that women should not have to become like men in order to be accorded the privileges of the male world. Instead, there should be greater recognition for women's values, cultures, and forms of behavior, all of which contribute to our nation's defense. This difference in opinion will continue to make the integration of women into the military a problematic process.

References

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

Bem, S. L. (1977). On the utility of alternate procedures for assessing psychological androgyny. Journal of Consulting and Clinical Psychology, 45, 196-205.

Cook, E. P. (1985). Psychological androgyny. New York: Pergamon Press.

DeFleur, L. B. & Warner, R. L. (1987). Air force academy graduates and nongraduates: Attitudes and self-concepts. Armed Forces and Society, 13, 517-533.

Devilbiss, M. C. (1985). Gender integration and unit deployment: A study of GI Jo. Armed Forces and Society, 11, 523-552.

Dunivin, K. O. (1988). Adapting to a man's world: United states air force female officers (Ph.D. dissertation). Evanston, IL: Northwestern University.

Hoffman, D. M. & Fidell, L. S. (1979). Characteristics of androgynous, undifferentiated, masculine, and feminine middle-class women. Sex Roles, 5, 765-780.

Hyde, J. S. and Phillis, D. E. (1979). Androgynous across the life span. Developmental Psychology, 15, 334-336.

Hyde, J. S. & Rosenberg, B. G. (1980). Half the human experience: The psychology of women, 2d ed. Lexington, MA: D.C. Heath and Co.

Moskos, C. C., Jr. (1985). Female GIs in the field. Society, 22, 28-33.

Segal, M. W. & Segal D. R. (1983). Social change and the participation of women in the american military. Research in Social Movements, Conflicts and Change, 5, 235-258.

Spence, J. T. & Helmreich, R. L. (1978). Masculinity and femininity: Their psychological dimensions, correlates, and antecedents. Austin, TX: University of Texas Press.

Tischler, H. L., Whitten, P. & Hunter, D. E. (1986). Introduction to Sociology, 2d ed. New York: Rinehart, Holt, and Winston.

Wood, F. R. (1982). U.S. air force junior officers: Changing professional identity and commitment (Ph.D. dissertation). Evanston, IL: Northwestern University.

Demographic and Attitudinal Correlates
of Women's Role in the Military

Michael D. Matthews
Drury College

Charles N. Weaver
St. Mary's University

Abstract

This paper presents data on American's attitudes toward women's roles in the military as a function of several variables. The data for this study come from the 1982 General Social Surveys, conducted by the National Opinion Research Center. The criterion variable required respondents to indicate whether women "should" or "should not" serve in the following military jobs: nurse in a combat zone, typist in the Pentagon, commander of a military installation, truck mechanic, jet transport pilot, jet fighter pilot, anti-aircraft gunner in the United States, hand-to-hand combat soldier, and crew member on a combat ship. Four predictor variables were examined: age, occupation, perceived relative financial standing, and education. In general, the results indicated that younger, more highly educated respondents who are employed in white collar jobs and perceive themselves to be above average in financial standing were signficantly more supportive of women's participation in all military roles examined. These effects were most pronounced for nontraditional jobs such as jet fighter pilot and hand-to-hand combat soldier. Implications for the future role of women in the all volunteer force or a draft-based force are discussed and policy ramifications examined.

During the recent invasion of Panama, Army Captain Linda Bray became the first woman to lead troops into modern combat. This action was widely reported by the media including an interview with Captain Bray on NBC's Today program, and coverage in major papers and news magazines. As pointed out in Time magazine (January 15, 1990, p. 29) the actions of Bray's platoon "rekindled a debate over whether women should be on the firing line."

One factor which bears on the integration of women into military roles concerns public attitudes toward such integration. If Americans express a favorable attitude toward women assuming combat roles, for example, political barriers to their full integration are lowered. Davis, Lauby, and Sheatsley (1983), examining national survey data which addressed this question, concluded that "Americans are not for total equality of the sexes in the military," but are in favor of a greater participation of women in the military than may be generally recognized. Matthews, Melton, & Weaver (1988) analyzed the same data base by Davis et al. (1983) and found that the attitudes of Americans toward women's roles in the military are significantly affected by the respondent's race and gender, with women being more supportive of women's participation in diverse and nontraditional roles than men, and blacks being generally less supportive of women's participation in any military role than whites.

The purpose of the current study was to further examine the attitudes of Americans toward women's roles in the military as a function of other demographic and attitudinal variables. More specifically, the current study examined support for women's participation in military roles as a function of age, education, perceived relative financial standing, and occupation. The data base that was used in the earlier study by Matthews et al. (1988) was also used in the current study.

Method

Responses from 1844 participants in the 1982 General Social Survey (GSS) conducted by the National Opinion Research Center provided the data base for the current study. The GSS involves face-to-face interviews of a national sample of United States citizens age 18 and over. Respondents are selected randomly, and the sample is stratified to insure representative sampling of major demographic groups. The GSS poses over 300 questions to each respondent on a wide variety of topics.

In the 1982 GSS, respondents were asked to indicate whether a woman "should" or "should not" be allowed to serve in the following military roles: nurse in combat zone, typist in the Pentagon, commander of a military installation, truck mechanic, jet transport pilot, jet fighter pilot, anti-aircraft gunner in the United States, hand-to-hand combat soldier, and crew member on a combat ship. These jobs provide a mix of traditional and nontraditional roles for women in the military.

Responses to these questions were crosstabulated with (1) the respondent's age (18-29, 30-49, and 50-89), occupation (white collar or blue collar), education level (less than high school, high school, or greater than high school), and perceived relative financial standing (lower than average, average, or higher than average). Estimates of statistical significance were made using Chi Square, and an alpha level of .05 was selected.

Results and Discussion

The main findings of the study are summarized in Table 1, which shows the percentage of respondents who indicated that women "should" be allowed to serve in each of nine military roles as a function of each of the predictor variables. With respect to respondent's education level, a clear trend toward favorable attitudes toward women's full participation in all military roles was found as education level rose. Respondents with less than a high school education uniformly showed the least support, ranging from 25.5 percent who indicated that women should serve in hand-to-hand combat roles, to 91.3 percent supporting women serving as typists in the Pentagon. Similarly, respondents with more than a high school education showed the strongest support, ranging from 41.9 percent who indicated women should serve in hand-to-hand combat, to 98.8 percent favoring women serving as typists in the Pentagon. All comparisons on this variable reached statistical significance.

Table 1
Proportion of respondents indicating a woman should be able to serve in various military roles as a function of respondent's education occupation, age, and perceived financal standing.

	Education (N=1844)			Occupation (N=1715)		Age (N = 1839)			Financial Standing (N=1827)		
	< HS	HS	> HS	white collar	blue collar	18-29	30-49	50-89	Below Average	Average	Above Average
Nurse in combat zone	86.2	92.1	94.7	93.9	88.5	91.5	92.3	88.9 NS	86.5	92.2	94.9
Typist in Pentagon	91.3	98.1	98.8	97.6	95.7	98.4	97.0	94.0	94.9	95.8	100.0
Military commander	40.6	60.7	70.5	63.9	51.3	64.1	66.1	41.7	53.0	54.9	67.3
Truck mechanic	65.8	85.0	90.7	87.5	75.5	87.6	83.1	72.2	75.6	80.7	87.9
Jet trans-port pilot	48.[illegible]	75.9	84.8	77.2	63.8	77.9	75.8	56.2	63.5	68.7	81.3
Fighter pilot	40.0	65.3	73.7	67.3	53.8	68.4	68.8	43.3	55.6	57.3	71.3
Air defense gunner in the U.S.	33.8	59.5	75.8	63.8	48.7	62.0	61.9	42.3	47.5	54.1	71.0
Hand-to-hand combat soldier	25.1	35.7	41.9	37.5	31.1	42.8	40.6	20.7	32.5	32.4	38.8 NS
crew member on combat ship	39.9	57.4	68.9	60.6	50.2	63.0	62.9	40.3	48.9	54.7	63.6

Substantial and consistent findings were also observed as a function of occupation. White collar workers clearly were more supportive of women serving in all roles examined. Again, all comparisons were statistically significant.

With respect to age, younger respondents proved to be more supportive of women's involvement in military roles than older people. As a rule, small differences were noted between respondents 18-29 and 30-49 years of age. However, in comparison to the responses of older (age 50-89) respondents, large differences were found, especially for nontraditional roles. For instance, only 20.7 respondents age 50 or older favored women's participation in hand-to-hand combat, compared to 42.8 and 40.6 percent for those ages 18-29 and 30-49, respectively. Among these comparisons, significant differences were found for all roles except for nurse in a combat zone.

Finally, the relationship between perceived relative financial standing and attitude toward women's roles in the military showed a consistent pattern of more favorable responses with increasing level of financial standing. This held true for each role considered except for hand-to-hand combat soldier, where no significant differences were found as a function of perceived financial standing. All other comparisons were statistically significant.

One aspect of the findings from this study which are not presented in Table 1 due to space restrictions is that very few respondents were unsure of their opinions on these questions. Of 99 comparisons made, in no case did more than 8 percent of the sample indicate "not sure," and in 92 of the comparisons the percentage of "not sure" responses was less than 5 percent. This held true even for nontraditional roles, such as hand-to-hand combat soldier, crew member on a combat ship, and jet fighter pilot. Apparently, most Americans have a firm opinion on these questions.

It is also interesting to note that the least support was given for the role of hand-to-hand combat soldier, with only 33.4 percent of respondents overall favoring women's participation in this role. The lack of support shown by the public for this role is surprising given that the majority of respondents favored women's participation in other combat and combat-related roles. For example, the overall percentage of respondents indicating that women should participate in the roles of crew member on combat ship, air defense gunner, and jet fighter pilot were 54.0, 54.2, and 58.7 respectively. For nurse in a combat zone and jet transport pilot, both of which are jobs that may be extremely dangerous in modern warfare, the overall approval rates were 90.8 and 68.7 percent, respectively. One wonders if Americans have an accurate view of the nature of the perils of modern warfare, given the responses of the subjects of this study.

One interpretation of the results is that Americans base their opinion on raw physical requirements of various roles, and in terms of direct contact with the enemy. Thus, the role of hand-to-hand combat soldier is viewed as inappropriate for women by a majority of Americans. In contrast, the role of fighter pilot, which may be equally dangerous, is viewed as more appropriate for women because of less direct (face-to-face) contact with the enemy and lesser physical demands placed on the pilot of modern aircraft. Consistent with this analysis, the role of typist in the Pentagon received the greatest approval by the respondents. It is also a low demanding job physically and is a great distance from enemy personnel.

One other dimension of the current data requires comment. The role of military commander received mixed support. Overall, 56.0 percent of respondents indicated that women should be in command of a military installation. This figure compares to those of some combat roles, such as fighter pilot. This is, of course, a nontraditional role for a woman in our society. Because support for this role was the same or smaller than other combat-related roles (e.g., nurse in a combat zone), it is possible that mere danger to the woman from the enemy does not explain all of the findings. One might hypothesize that the greater the deviation of a role from cultural stereotypes, the less the support for it. So, protective paternalism may not be the only factor in interpreting the results found in this study.

It is clear from the current data that a number of variables modified the respondent's attitudes towards women's roles in the military. Education, age, occupation and perceived financial standing (and presumably other variables) affect how one views this issue. Information of this sort shoudl be of value to policy makers and planners in building support for changes in personnel policy relating to women's roles in the military. They also have implications for recruiting and human resource projections. If, for example, the applicant pool for the Army is heavily represented by poor, relatively uneducated persons from blue collar backgrounds, then a change in Army policy regarding utilization of women in combat roles might impact recruiting. A woman interested in joining the Army might decide not to apply if she felt likely to be placed into a role that she and her friends and family feel is inappropriate for women. The possible effects of policy changes on recruitment need to be examined systematically.

Returning to the case of Captain Bray, one would conclude from the current study that the majority of Americans favor an expansion of military roles open to her and other women, but are not at this point ready to accept women in hand-to-hand combat actions. The dynamics behind these attitudes are worthy of in-depth and systematic examination, and the ramifications of future policy changes must be explored in light of effects they may have on other military programs and operations.

References

Davis, J. A., & Smith, T. (1987) General Social Surveys, 1972-1987: Cumulative Codebook. Chicago: National Opinion Research Center.

Davis, J. A., Lauby, J., & Sheatsley, P. B. (1983). Americans view the military. (Tech report/NORC Rep. No. 131). Chicago: National Opinion Research Center.

Matthews, M. D., Melton, E. C., & Weaver, C. N. (1988). Attitudes toward women's roles in the military as a function of gender and race. Proceedings of Psychology in the Department of Defense, Eleventh Symposium. (USAFA TR-88-1). Colorado Springs, Colorado: United States Air Force Academy.

Time Magazine (January 15, 1990). Fire when ready, Ma'am. Vol. 135 (3), p. 29

Modifying the USAFA BCT Program to Reduce Gender Discrimination Among Basic Cadets

Cynthia A. Wright

"In leadership, there appears to be a continuing relationship between physical performance and leadership ratings for women...Women who do not perform well on physical performance measures typically receive lower ratings in leadership by the men."

--Maj Jerome Adams, 1979

Abstract

The Basic Cadet Training program (BCT) is designed to teach military skills and discipline, strengthen leadership and followership skills, develop physical fitness, and foster a spirit of teamwork and comraderie while enhancing self-confidence. In order to accomplish all of these objectives, the program must be physically strenuous as well as mentally challenging, and yet be well within the abilities of most individuals so that all can have a reasonable expectation of being able to excel through their own efforts. As it is presently administered, the USAFA BCT program seriously hinders the development of leadership skills, teamwork, and self-confidence in female basic cadets because of its overemphasis on specific physical activities which do not allow recognition of overall physical conditioning or mental aptitude. Modifications of the BCT conditioning and intramural programs and the addition of more non-physical evaluations and activities are recommended in order to lessen the gender bias that currently exists at USAFA.

When, on the first day of basic training, cadets find themselves rather dramatically separated from their civilian lives and relationships, they immediately begin to replace those ties by establishing their social place in the unfamiliar environment. Their evaluations of one another in the first couple of weeks have a considerable effect on their lasting impressions of relative leadership potential and of each individual's ability to contribute to the effectiveness of the unit as a whole. Although much of the training is necessarily physical, it tends to focus very narrowly on those activities in which most men, with their generally greater size and upper-body strength, are inherently more adept than the majority of women. As indicated in Adams' statement above, based on his observations of male and female cadets at West Point, the females' perceived leadership abilities are considerably decreased by their general inability to perform as well as their male peers on these few exercises, especially running.

The emphasis on skills in which females typically do not excel creates a general image of women as liabilities, rather than assets, to any group or team. According to Priest, Prince, Rhone, and Vitters (1977), "Relatively few women, not more than 7%, exceed the male thirty-third percentile standard. This means there is little overlap in height, weight, or physical aptitude measures of the genders." Despite the fact that an average woman's inability to match a particular physical feat performed by an average man is not necessarily an indication of her relative fitness, "a large minority of men...want equal treatment with no exceptions" (Priest, et al., 1977). This attitude fosters the emergence of an informal leadership heirarchy based almost exclusively on physical aptitude in activities at which men typically excel.

Greeno (1989) argues that, in education, individuals' personal epistemologies reflect norms (such as the bases for determining leadership potential). These normative perceptions in turn profoundly affect group members' aquisition of skills. The establishment of leadership roles during BCT is based partially on the traditional gender roles predominent in the backgrounds of many cadets, especially males-- women at USAFA are less likely to reflect traditional sex-role socialization, since attending a military academy is generally considered to be inconsistent with expected "feminine" behavior. The insistence on physical performance as the primary indicator of overall leadership ability creates an environment in which men will naturally be more likely than women to achieve traditionally masculine leadership roles. Women who conform to the imposed norms out of a need to belong to the group are less inclined to seek out opportunities for leadership because doing so would force them to act contrary to the expectations of their predominently-male peers.

The male basics' initial perceptions of women as generally inferior because of their physical aptitude often carry over inappropriately to the fourthclass year, in which athletic performance should be largely replaced by academic and military performance as the criteria for leadership. This perceptual set, based on first impressions, may prevent female cadets from being able to adequately establish themselves as equally contributive members of their unit even if they excel in all non-physical aspects of training during the academic year. The lack of recognition for their leadership abilities as measured by academic or military performance is damaging not only to the women's self-esteem, but also to the overall effectiveness and comraderie of the group or squadron as well.

In order to provide for a greater equality of treatment both during and after BCT, modifications should be made to the training program which allow cadets to compete in non-gender-specific activities. The first recommended change is the elimination of the morning physical conditioning (PC) run. According to Adams (1979), "Women continue to fall out of the runs in greater numbers than do the men in Cadet Basic Training." This tendency still manifests itself today. Not only are most female cadets at a severe disadvantage in an activity which is largely affected by height (and corresponding length of leg), but more females suffer running-related injuries than do men because of the additional stress placed on knees and ankles by wider hips. Females comprised only 10% to 12.5% percent of their respective incoming classes from 1987 to 1989, yet statistics from the USAFA Cadet Clinic indicate that 32.6% of the leg, knee, and ankle injuries reported by basic cadets during those years were experienced by women.

Another area in which modifications could be made is daily intramurals. Since BCT intramurals are designed to promote leadership in a team format as well as promote physical conditioning, the program should be reevaluated in terms of creating equal opportunities for women, and men with less athletic prowess, to excel. Priest (1977) suggests that "intramural competition could serve such a purpose [developing more positive conditions for intergroup contact] if it were organized to emphasize skills in which women are the equal of men." Since there are few athletic events in which women and men are equally proficient, one solution would be the implementation of a rotating schedule in which elements or flights would participate in a different sport every week. Currently, intramural teams are chosen at the beginning of BCT for such team sports as flickerball and basketball. These teams remain constant, and compete in their sport every afternoon during the first three weeks of BCT. A rotating schedule would allow the elements or flights to participate, either as a team or simply a group, in several different activities, such as the usual team sports, aerobic workouts, cross country, swimming, etc.

Although men would continue to excel in certain areas, the addition of activities in which women are traditionally more adept (such as aerobics), and in which neither gender necessarily has an advantage (such as swimming or volleyball), would allow recognition of more diverse physical skills than those allowed by the present program. This has the double advantage of promoting teamwork among the group, since elements or flights would not be split among different teams, and of continuing to allow individual leaders to emerge in each athletic endeavor.

Adams' (1979) recommendations suggest a final modification to the physical training environment, focused on the cadre, rather than on the basic cadets themselves:

> More time needs to be set aside in the cadet cadre preparation training to allow the cadets to learn more about the physiological differences which account for physical performance differences between men and women during summer training...The leadership program should devote one session to explain that physical

performance, especially run exercises, is not the only measure of a cadet's ability to lead.

In order to deemphasize the powerful informal norm which recognizes only physical standards of leadership, a corresponding increase in mental and military evaluations should be implemented. Priest, et al. (1977), point out that "women cadets tend to have achieved greater rank in high school academic subjects than men cadets. Without a "job analysis" which justifies a preference for certain traits in...graduates, there is no logical basis for preferring physical aptitude over mental aptitude in a candidate." The generally greater military and academic achievements by women are substantiated by the USAFA Office of Institutional Research (1988), which notes that there has consistently been a greater percentage of firstclass women than men on the Dean's, Commandant's, and Superintendent's lists. With this evidence of non-athletic aptitudes in mind, another possible modification would be the restructuring of BCT Field Day to more closely correspond to the Sweepstakes event in Recognition training at the end of the fourthclass year. Whereas Field Day incorporates only athletic contests, Sweepstakes focuses on the accomplishment by small groups of various "obstacles," which may focus on either physical challenges or military knowledge. A combination of the two measures would result in a day of general competition in sports, BCT knowledge from Contrails, and military drill, all of which are significant parts of basic training. Such squadron competition would place equal responsibility for winning or losing upon those who excel physically, mentally, and militarily.

A means of discovering and developing the various non-physical leaders prior to the actual Field Day competition might include constant competitions within the squadrons in every area of military training. The top performers in each area, such as memorizing quotes and knowledge, shoe-shining, bed-making, or rifle drill would train several of their peers who are moderately skilled in the same area. Each team would then be responsible for tutoring the rest of their classmates in their area of specialization during element leader blocks. Not only would such a program increase the overall skill levels within the squadron, but it would also help increase teamwork, develop leadership skills, and promote class unity in terms of the entire group being responsible for each member.

As the nature of warfare continues to shift toward reliance on technological weaponry, rather than the traditional hand-to-hand combat, the potential for significant contributions by women, especially as members of heterogeneous problem-solving groups, will increase dramatically. If sexism is to be reduced throughout the cadet wing and the Air Force in order to best utilize women as a vital resource, the process must start with the BCT cadre and with the basic cadets. Because many cadets, especially males, come from families with traditional values in terms of gender roles, and because their initial impressions of their peers frequently affect future evaluations of individual leadership potential--particularly that of women-- basic cadets' initial training should stress non-gender specific aptitudes in as many different areas as possible during their indoctrination into a military career.

REFERENCES

Adams, Jerome. Report on the Admission of Women to the U.S. Military Academy: Project Athena III. Washington D.C.: Defense Logistics Agency, June 1979.

Greeno, James G. "A Perspective on Thinking." American Psychologist. American Psychological Association, Inc., 1989.

Priest, Robert F. The Intergroup Hypothesis as Applied to Women at West Point. Washington D.C.: Defense Documents Center, June 1977.

Priest, Robert F., Prince, Howard T., Rhone, Teresa, and Vitters, Alan G. Differences Between Characteristics of Men and Women New Cadets, Class of 1980. Washington D.C.: Defense Documentation Center, March 1977.

USAFA Office of Institutional Research. Performance of Women in the classes of 1980-1991 (Table 2). USAFA, 1 JAN 88.

USAFA Cadet Clinic Records. Computer printout on BCT Injuries 1987-1989. Obtained from Msgt. Vance Marquis, December, 1989.

Motion Sickness Prevention: A Course of Instruction in Cognitive-Behavioral Counseling[1]

Thomas G. Dobie
Naval Biodynamics Laboratory
James G. May
University of New Orleans

Abstract

This paper concerns the evaluation of a program of instruction aimed at training individuals to counsel clients who are prone to motion sickness. Unselected volunteers were offered a course of instruction reviewing the cognitive-behavioral approach to help personnel tolerate motion environments without developing motion sickness. The methods of presentation, content and assessment of knowledge gained are described. Eleven participants of that course were subsequently available to independently counsel clients who were prone to motion sickness, using the cognitive-behavioral approach with the aid of reinforcement by visually-induced apparent motion. The clients were pre- and posttested by an independent observer using tolerance and motion response as the dependent variables. These test scores were compared to previous data obtained with clients who had received counseling from an experienced counselor, or had received no such counseling. The results indicated that the newly trained counselor's clients showed significantly increased pre- to posttest tolerance to the motion stimulus, although they did not benefit as much as clients trained by the experienced counselor. In addition, however, the trainee's clients exhibited as much benefit, in terms of posttest symptomatology as did those of the experienced counselor. These data are taken as strong support for the feasibility of training counselors to employ this method of alleviating motion sickness.

The term motion sickness refers to a cluster of signs and symptoms observed in individuals during or shortly after experiences in some motion environments. The etiology of this syndrome has been the focus of numerous reports (Money, 1970; Reason & Brand, 1975; Muir, 1983; Benson, 1984) and is generally considered to be associated with disorientation which results from the inability of the human sensory apparatus to contend with motion environments provided by modern transportation. More specifically, it is felt that motion sickness comes about because of a mismatch in the sensory inputs during motion (Benson, 1984; Oman, 1982; Reason, 1978; Reason & Brand, 1975). Any disruption of the natural correlation of these inputs, or the internal dynamic model, results in the syndrome of autonomic responses commonly described as motion sickness.

---

[1] The opinions and interpretations contained herein are those of the authors and do not necessarily reflect the views, policies, or endorsements of the Department of the Navy or any other government agency.

The incidence of motion sickness is difficult to determine, since it depends on the strength, characteristics, duration of motion stimuli and the individual's previous experiences with motion. Some vestibular physiologists contend that motion sickness is reflexive and occurs in all normal clients given a provocative enough motion environment. But it is also clear that some individuals are more susceptible to the effects of motion than others. One recent view suggests that some of the individual differences derive from attitudes and expectations about motion environments formed from previous experiences in motion laden settings. It is felt that anticipatory fears regarding motion environments serve to exacerbate the reflexive response to disorienting motion and render the individual more prone to motion sickness. If this view is correct, it suggests that one approach to the treatment of motion sickness might include cognitive intervention.

In a recent series of investigations (Dobie et al.,1987; Dobie et al., 1989), Dobie and coworkers have investigated various forms of treatment aimed at increasing tolerance to motion or apparent motion experiences. The results of these studies indicated that the most beneficial treatment involved counseling which sought to increase an individual's confidence in their ability to tolerate disorientation during motion, coupled with controlled exposure to such motion to reinforce those beliefs. Since previous experiments have involved only one counselor, a concern has been whether or not these treatment procedures are efficacious in the hands of other individuals. The focus of the present study was to develop a course of instruction which provides counseling expertise and to field-test the results of such instruction with clients prone to motion sickness.

Method

Trainees

Thirteen individuals, ten civilians and three Navy personnel were enrolled in a 12 session course of instruction on motion sickness prevention. This included information on the signs and symptoms of motion sickness, etiology, treatment, and prevention. They were asked to read Benson's review article and a series of articles concerned with the cognitive-behavioral method of training tolerance to motion environments (see references). This approach points out that in addition to the normal disorientation and anxiety experienced in some motion environments, the individual approaches such activities with attitudes, memories and past experiences associated with similar motion environments. These anticipatory fear reactions can exacerbate or amplify arousal leading to increased susceptibility to motion sickness. The aim of cognitive-behavioral therapy is to help the client understand this and to build confidence in the belief that alleviating the negative cognitive overlay will allow them to tolerate motion environments. Such confidence can be strengthened by controlled, incremental, exposures to disorienting motion or pseudo-motion environments. The theoretical and experimental support for the feasibility of this method was discussed and specific case histories were reviewed. A fifty item multiple choice examination was administered in the last session. Grades ranged from 66 to 96 % correct with a mean of 85.5 % and a standard deviation of 7.7 %.

Clients

All potential clients were asked to fill out a motion history questionnaire and a medical history questionnaire On the basis of this information, clients with a history of motion sickness and no medical problems were asked to undergo pretesting as described below. Those clients who's pretest tolerance score was less than 8 min. were selected and assigned to counselors. They ranged in age from 17 to 53 years; 2 were male and 18 were female. Clients were always assigned to a counselor whom they had not previously met. Nine of the eleven counselors were assigned two clients and the two remaining counselors were assigned only one. Thus, twenty clients were treated by the eleven counselors.

Apparatus

Tolerance to visually-induced apparent motion was used as a dependent measure and as a means to reinforce client's confidence regarding their ability to tolerate disorientation. The device was a circular drum five feet in diameter and four feet in height with a mirrored ceiling. The inner surface of the drum was lined with alternating black and white vertical stripes, six inches wide. Clients sat in a stationary chair while the drum rotated around them at 10 rpm. With appropriate fixation the entire visual field was stimulated and rotation of the drum produced a compelling illusion of circular self-motion.

Procedure

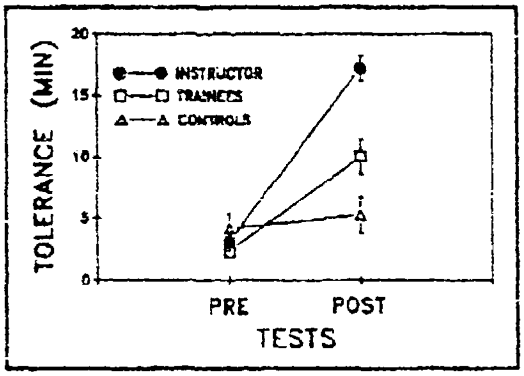
Prior to treatment, all clients completed a current physical status report and, before and after stimulation, a symptom checklist. During pretest, clients were instructed to indicate, using a button press, when they first noticed identifiable motion sickness responses, at which time the stimulation would cease. Twenty minutes was set as the limit for tolerance. All clients were pre- and posttested by the same experimenter who was not one of the counselor candidates. The posttest was identical to the pretest. All counselors were instructed to employ cognitive-behavioral counseling over ten sessions. They were invited to seek clarification, from the experienced counselor or the principal investigator, if they had any questions as to the procedures to be employed. No assistance, however, would be provided during a counseling session.

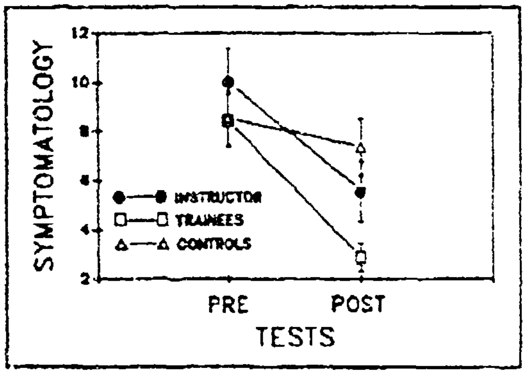
RESULTS

Measures of pre- and post treatment tolerance and symptomatology were tabled for the twenty clients in the present study, together with eighteen clients previously treated by the experienced counselor and eleven non-treated control clients. Each measure was submitted to a two factor, mixed analysis of variance design. Subsequent paired comparisons were performed with Newman-Keuls' tests. Significance levels were set at $p < 0.01$ for all effects.

The mean tolerance scores as a function of pre-posttesting for all three groups of clients are presented in Figure 1. It is apparent that the mean tolerance of both the instructor's and the trainee's clients increased considerably relative to the mean tolerance of the control group. However, the instructor's clients showed greater benefits than those of the trainees.

Significant main effects for groups and prepost were obtained. A significant groups by prepost interaction was also obtained. The pretest tolerance scores were not significantly different, nor were the pre- and posttest scores for the control group. All other paired comparisons were significant.





Figures 1 and 2. Mean tolerance (left) and symptomatology (right) scores as a function of pre- and posttesting for clients treated by instructor and trainees, together with untreated control subjects.

The mean symptomatology score was obtained by summing the difference in symptoms before and after each pre- and posttest for each individual, to obtain a composite score for each test. The mean symptomatology scores as a function of pre- and posttesting for all three groups are presented in Figure 2. It is apparent that all three groups exhibited a decline in symptoms, but the treated groups showed a greater decline than did the controls. A significant main effect for prepost, but not for groups was found. A significant groups by prepost interaction was also obtained. The mean pretest symptomatology scores were not significantly different, nor were the pre- and posttest means for the control group. Significant pre-post mean differences were obtained for each treated group, and the trainee group's posttest was significantly below that of the other two groups. On posttest the instructor group's mean was not significantly different from the control (probably due to a poor sampling break obtained at pretest).

DISCUSSION

These results clearly support the feasibility of the counseling program and offer further support for cognitive-behavioral training as a viable method to manage motion sickness. Although these data are impressive, numerous factors should be considered in the future implementation of counselor training. In the present investigation, we made no attempt to select trainees according to any preset criteria. Thus, our training program was a conservative test of the notion that such skills can be taught. It could be expected that the inclusion of numerous selection criteria would lead to even more efficacious training. It might be useful to explore various standardized tests of personality, aptitude, ability and achievement as selection criteria. It is our impression that an assessment of potential trainee's confidence, interpersonal skill, maturity, intelligence, persuasiveness and optimism are all important. While standardized tests of such ability may not be available, it is feasible to assume that these characteristics can be assessed by interview.

While the present course content and method of instruction appear adequate, some modifications might prove beneficial. Some trainees suggested that it might be helpful to view video tapes from actual counselor sessions. We recommend that this be considered for inclusion. We used a conventional lecture format, but it might be that the material could be covered equally well with program instructional media. Thus, self-study programs in the form of programed texts or computer based, interactive programs should be considered. It may be that group discussions lead by an experienced counselor would be a useful adjunct to any mode of instruction. Finally, it is highly recommended that a period of supervised counseling be added to the future implementation of these training efforts.

In the present investigation, we used visually-induced apparent motion exposure during counseling to reinforce subject confidence in their ability to cope with disorienting environments, but the apparatus to produce such experience may not be readily available. We are presently exploring the possibility of using other modes of motion to produce disorientation (e.g. active bodily rotation), which may prove more economical. In conclusion, it is apparent that considerable counseling skill can be imparted to inexperienced trainees and and these individuals can help personnel learn to tolerate provocative motion stimuli. This has important implications for both military and civilian populations - on the ground, in the air or at sea.

REFERENCES

Benson, A. J. (1984). Motion Sickness. In M. R. Dix, & J. D. Hood, (Eds.), Vertigo. Chichester, England: John Wiley & Sons.

Dobie, T. G., May, J. G., Fisher, W. D., Elder, S. T., & Kubitz, K. A. (1987). A comparison of two methods of training resistance to visually-induced motion sickness. Aviation, Space and Environmental Medicine, 58, A31-41.

Dobie, T. G., May, J. G., Fisher, W. D., & Bologna, N. B. (1989). An evaluation of cognitive-behavioral therapy for training resistance to visually-induced motion sickness. Aviation, Space and Environmental Medicine, 60, 307-314.

Money, K. E. (1970). Motion sickness. Physiological Review, 50,1-39.

Muir, C. A. (1983) Motion illness - a bibliography. London: Medical Research Council, Royal Naval Personnel Research Committee, Ship Motion Working Party report, SMWP 1/83.

Oman, C. M. (1982). A heuristic mathematical model for the dynamics of sensory conflict and motion sickness. Acta Otolaryngologica, Supp. 392.

Reason, J. T. (1978). Motion sickness adaptation: a neural mismatch model. Journal of the Royal Society of Medicine, 71, 819-829.

Reason, J. T. & Brand, J. J. (1975) Motion Sickness. London: Academic Press.

Pilot Briefings and Aircrew Coordination

Evaluation: Empirical Results

Thomas M. Franz
G. Andre' McCallum
Michele D. Lewis
Oak Ridge Associated Universities
Carolyn Prince
Eduardo Salas
Naval Training Systems Center

Abstract

This paper outlines an investigation of the relationship of behaviors exhibited in a pre-flight brief with subsequent performance in a mission simulator. Seven helicopter crews were evaluated on their Mission Analysis, Assertiveness, Flexibility, Situational Awareness, Leadership, Decision Making, and Communication, during the preflight brief and subsequent flight. Operational behavior ratings were also made for the simulated flight. Leadership pre-flight briefing behaviors were found to be related to Operational ratings and to the Leadership behaviors observed during simulated flight. Decision Making behaviors displayed during the pre-flight brief were positively correlated with Decision Making behaviors exhibited during simulated flight.

Over a decade ago, researchers reported that a large number of aircraft incidents and accidents were due to a lack of decision-making, leadership, pilot judgement, and coordination in the cockpit (Cooper, White, & Lauber, 1980). Since that time civilian and military aviation experts have worked to design and implement aircrew coordination training (ACT) programs in hopes of avoiding future accidents and incidents. However, in 1987, Hackman noted that the parameters in which the crew must operate are largely established prior to the crew's first meeting, and that many ACT programs ignore this first (and possibly most crucial) interaction. Ginnett (1987) reported that during the first few minutes of the pilot's pre-flight brief, crew members can reliably assess whether the pilot will be effective at building and promoting a positive working environment or not.

Ginnett's (1987) field study of commercial pilots who were judged by their peers to be either effective or ineffective in crew coordination revealed specific pre-flight briefing behaviors that were exhibited by effective pilots. A listing of some of the behaviors identified by Ginnett is given on the left side of Table 1. On the right side of Table 1 are behavioral examples of effective military coordination (Oser, McCallum, Salas, & Morgan, 1989). As seen in Table 1, although the applications differed between civilian and military teams, there were many similar behaviors identified by these independently conducted research efforts. The Oser et. al. behaviors, along

Table 1. Contrast of Ginnett's (1987) Briefing Behaviors and Oser, et al., Military Coordination Behaviors

Ginnett's Briefing Behaviors	Oser, et al., Behaviors
Engage Others in Conversation	Reply with Question/Comment
Created Boundaries	Structure Tasks, Plans, Objectives
Provide Rationale	Provide Rationale
Seek to Understand/Not Place Blame	Provide Feedback on Performance
Engage Others in Brief	Ask For Input
Discuss Coordination Tasks	Verbalize Plans for Procedures
Disavows Perfection	Provide Legitimate Avenue of Dissent
Tailor Brief to Mission	Question, Seek Data, Information, Plans

with other behaviors identified as critical to aircrew coordination performance, served as a basis for the behavioral skills rating instrument used in this study (Franz, Prince, Cannon-Bowers, and Salas, 1990).

The purpose of this investigation was to assess the relationship of pre-flight briefing behaviors with subsequent performance in a full motion helicopter flight simulator. The pre-flight brief served as the formation and orientation period for each crew. A written outline for the brief was followed by every pilot, as required by regulations. This period allowed the pilot to disseminate relevant information (e. g., weather, ceilings, landing zone coordinates) and to verbalize procedures (e.g., transfer of flying responsibilities, emergency procedures, etc.) that may be required during the mission. It was expected that observations of behavioral examples like those in Table 1, during the preflight brief, would be associated with observed aircrew coordination behaviors and operational ratings in a simulated mission.

Method

Each preflight brief was tape recorded, each crew's simulator performance was video taped, and the tapes were subsequently transcribed. The raters' scores were based on these transcripts. Raters thoroughly reviewed the transcripts, and then independently indicated which behavior(s) were displayed by the crew. The rater then assigned a numerical score for each crew member on each dimension (0 = Not Observed, 1 = Unsatisfactory, 2 = Below Average, 3 = Average, 4 = Above Average and 5= Excellent). The preflight briefs were rated on seven behavioral dimensions: Decision Making, Assertiveness, Flexibility, Situational Awareness, Mission Analysis, Leadership, and Communication. A coefficient of inter-rater agreement was calculated to quantify the amount of agreement between raters. A coefficient for overall agreement was calculated at ($r = .76$).

The transcripts of the simulator scenarios were rated on the same seven dimensions. A safety officer qualified to rate the crews' operational performance also gave operational ratings that were used as a dependent measure. Correlations were then computed between the ratings to determine whether preflight brief behaviors were associated with aircrew coordination performance and/or operational performance.

Results

As seen in Table 2, the Leadership behaviors were most highly correlated with the operational performance ratings. That is, crews that addressed topic areas related to Leadership during the preflight brief received higher ratings of overall task performance. Frequent behavioral observations within this dimension included, but were not limited to: the pilot asking for input and discussing potential problems, the pilot informing crew members what should be done, and the pilot providing the crew members with a legitimate avenue of dissent to voice their concerns.

Correlations of the preflight briefings with the Pilot and Co-pilots behavioral ratings in the full-mission simulator facilitated an examination of the consistency of the crews' behaviors from the brief to the mission. As indicated in Table 2, Leadership and Decision Making were positively correlated on theses two measures. Leadership behaviors in both the preflight brief and the scenario included determining clearly the tasks to be assigned, and keeping the crew focused on the

Table 2. Correlations (r) of Pre-Flight Briefing Behaviors
with Operational Ratings and Crew Coordination Behaviors

Pre-Briefing Dimension	Operational Ratings	Pilot & Co-Pilot Behaviors	Pilot Sim Behaviors	Co-Pilot Sim Behaviors
Mission Analysis	.37	---	---	---
Decision Making	-.14	.97*	.77*	---
Assertiveness	.48	-.08	-.33	.94*
Adaptability Flexibility	---	---	---	---
Situational Awareness	.21	-.32	-.32	---
Leadership	.79*	.91*	---	---
Communication	.12	-.09	---	-.19

Note: N=14. * $p < .05$. --- insufficient data task at hand. Specific Decision Making behaviors consisted of members gathering information before making a decision, crew members cross checking information, and members frequently identifying alternatives and contingencies available to them.

The last two columns in Table 2 provide insight to the specific behavioral dimensions relating to the preflight briefing performance of pilots and co-pilots. The extent to which a pilot will display effective decision making behaviors during a flight was associated with his rating on the decision making dimension during the preflight brief. Pilots who initiate effective decision making strategies, tended to use effective decision making strategies during the mission. The behaviors consistent for the co-pilot during the preflight brief and during the mission lay in the assertiveness dimension. Co-pilots who frequently asked questions when uncertain, made suggestions, stated opinions and confronted ambiguities during the preflight brief were also assertive during the mission.

Discussion

Previous independent research (Ginnett, 1987; Oser, McCallum, Salas, & Morgan, 1989) has revealed comparable behaviors commonly demonstrated by effective crews during their initial formation. In the present research, the results indicated that a qualified military aircrew's behavior, as observed during the preflight brief, was associated with their operational performance ratings and their coordination levels in the cockpit.

The findings were supportive of earlier work in the civilian aviation sector, and offer promise for future research of the aircrew preflight briefing process. As expected, the crews displayed variations in preflight briefing behaviors despite following a standard checklist . In summary these preliminary results suggest that several relationships exist between the preflight brief and performance in the cockpit. Preflight brief behaviors were related to: (a) operational crew performance, (b) an aircrew's effectiveness as measured by the dimensions of Leadership and Decision Making. (c) a pilot's decision making effectiveness, and (d) the extent to which a co-pilot will display effective assertive behaviors in the cockpit.

References

Cooper, G. E., White, M. D. & Lauber, J. K., Eds., (1980). Resource management on the flight deck. NASA Conference Publication 2120.

Ginnett, R. C. (1988). Cockpit crew effectiveness from the inside out: A micro-analysis leading to macro-considerations. Proceedings of the Eleventh Psychology in the Department of Defense Symposium.

Hackman, R. J. (1987). Group level issues in the design and training issues of cockpit crews. Cockpit Resource Management Training, (Tech. Report No. NASA CP-2455). Moffett Field, CA: NASA Ames Research Center.

Oser, R., McCallum, G. E., Salas, E. & Morgan, B. B. Jr. (1989). Toward a Definition of Teamwork: An Analysis of Critical Behaviors. (Technical Report No. NTSC 89-004). Orlando, Florida. Naval Training Systems Center.

The Identification of Aircrew

Coordination Skills

Thomas M. Franz
Oak Ridge Associated Universities
Carolyn Prince
Janis A. Cannon-Bowers
Eduardo Salas
Naval Training Systems Center

Abstract

This paper gives a brief outline of the methodology used to identify and categorize behaviors used as a basis for development of a skill-based aircrew coordination training program for a helicopter community. The research effort began in 1986 and encompasses literature reviews, observations of team development, interviews of job experts, and surveys of the aviation community. Thirty seven behaviors were identified and categorized under the dimensions: Mission Analysis, Decision Making, Assertiveness, Flexibility, Situational Awareness, Communications, and Leadership.

The fact that aircraft incidents and accidents today are largely caused by human error is compelling justification for the introduction of aircrew coordination training programs. While early aviators were plagued with accidents caused by faulty equipment or a lack of "stick and rudder skills" it is human error that is the identified cause of 60-80% of all incidents and accidents today (Carroll and Taggart, 1987). These human errors are not due to a lack of knowledge needed to fly the aircraft, but from failures to coordinate the resources available to the pilot, particularly, the resources of the fellow crewmembers (Lauber, 1987).

Poor crew coordination compromises not only flight safety but also effective mission performance. Confusion about role requirements in non-routine situations, the failure to include all crew members in decisions that require their inputs, and a lack of awareness of the work and information needs of fellow crew members can reduce the effectiveness of crew performance. These two factors, safety and performance, have spurred a new emphasis on training crews for optimal coordination in the cockpit.

Although crew coordination training is important for all aircrews, it is of particular importance for military aviators. The nature of military missions, with their potential for increased stress and workload, makes coordinated action in the cockpit even more critical than in routine civilian flying. Military aviator training must be designed to assure the acquisition of the coordination skills needed for both present and future mission requirements. In order to assure that the necessary training occurs, the identification of these coordination skills is required. Once these skills are identified, qualitative and quantitative measurements of their occurrence can then be used to validate the use of a skill-based training program.

The aviation industry and the military have developed programs that attempt to meet aircrew coordination training needs. However, these programs vary in their approaches, duration of training, in identification and definition of necessary skills, and training technology used (Jensen, 1987). Diversity in the training is due in part to the lack of validation of the programs. Despite the importance of the training and the widespread interest in the programs, both by those in air operations and the Federal Aviation Administration (FAA), their effectiveness in improving cockpit coordination has not been established (Jensen, 1987; Foushee, 1987).

It is clear at present that: (1) the problem of human interaction for the safe conduct of flight is considered serious enough that it must be addressed, and (2) the elements that need to be trained, as well as the most effective technology to be used, have not yet been determined. It is the identification, development, and evaluation of behaviorally based crew coordination skills that is considered primary in aircrew coordination training programs. The focus of skill-based aircrew coordination training is behavior. No matter what the theoretical base for the existing programs, most of them refer to behaviorally-based skills, which are considered important to effective cockpit management. One researcher has pioneered efforts in this area. Ginnett (1988) observed airline captains who exhibited dramatic behavioral differences in their ability to build and maintain teams. Ginnett's observations of behavioral differences during the briefing of the crew, suggested that these behaviors can be identified, taught, modeled and evaluated.

Part of the failure of existing aircrew coordination training programs to evaluate training effectiveness is the absence of a means for measuring the effects of the training. It is important that aircrew coordination training include observable elements that will facilitate evaluation. This paper outlines the process of identifying behaviorally based crew coordination skills for a military helicopter community.

Research Foundation

A review conducted by Morgan, Glickman, Woodward, Blaiwes and Salas (1986) revealed a listing of behaviors that were expected to occur during the formation/evolution of the crew. Morgan et al., suggested that coordination training should seek to improve and enhance the crews' ability to communicate, relate, and interact. Furthermore the training should "generate group cohesion and organizational commitment, and sustain the integrity and viability of the team" regardless of the situational circumstances (p. 17-18).

In 1986, Morgan et al., verified that these behaviors or skills are needed for the development of effectively functioning crews. Using a collection of critical behaviors from instructors and self reports from team members, Morgan et al., identified ninety behaviors characteristic of crew development. The behaviors were categorized into seven dimensions: communication, coordination, team spirit and morale, cooperation, adaptability, acceptance of suggestions or criticism, and giving suggestions or criticism. Morgan et al's., observations of Naval Gunnery Crews revealed that these dimensions could be used to discriminate between effective and ineffective crews. Effective crews exhibited a higher number of critical behaviors in each dimension, and their process of crew development could be observed using the identified behaviors.

According to Morgan et al., crew coordination training needs to be committed to those person-to-person activities which are designed to enhance interpersonal communications, social relationships, and interaction patterns (i.e. the maintenance of the crew as a cohesive unit). As Morgan et al., hypothesized (and Ginnett found) a "substantial portion of the energies devoted to building better crews can be accounted for in terms of activities that are aimed at people (i.e., other team members) and relationships" (p. 17).

In 1989, Oser, McCallum, Salas, and Morgan concluded that (a) it was possible to identify behaviors frequently used by crews, (b) discriminations can be made between more and less effective crews based on behavioral observations, (c) existing behavioral profiles are enhanced with the addition of behavioral examples, (d) the behavioral examples added to our understanding of team development, (e) these behaviors can predict team success, and (f) these behaviors can be used to define and identify the dimensions: Communication, Cooperation, Team Spirit and Morale, Giving and Accepting Suggestions or Criticism, Coordination and Adaptability. Subsequent research revealed how these "generic" behaviors can be tailored to fit an existing helicopter community.

Method

The approach taken to accomplish the objective of determining which team building behaviors were related to aircrew coordination was as follows. A review of past aircrew coordination literature and team training literature revealed 37 behavioral statements related to aircrew coordination. Interviews of 20 pilots in the targeted community revealed an additional 18 behaviors which were added to the list. Twenty one job experts were then asked to verify the importance, criticality, difficulty, and frequency of occurrence of these behaviors on 7-point scales (where 7 was the high end of the scale) via survey responses.

Based on the responses to the first survey, revisions were made to reduce ambiguity of items and to insure that the items "made sense" to the pilots. The survey was then readministered to a second sample of 134 job experts, and it was found that overall, the aircrew coordination behaviors were considered very important to train in pilots ($\underline{M}$ = 5.5) and were very critical to mission effectiveness ($\underline{M}$ = 5.7). Frequency of occurrence was somewhat lower ($\underline{M}$ = 4.2) and difficulty was below the midpoint of the scale ($\underline{M}$ = 3.7). The level of agreement appeared to have been satisfactory as evidenced by the standard deviations of the item ratings (78% were less than a 1.5 standard deviation).

Correlations between the mean criticality, difficulty, frequency, and importance to train scores were then calculated. As is evident in Table 1, difficulty showed somewhat lower association than other rating categories, indicating that subjects did not necessarily consider the behaviors to be difficult, but that they are critical and should be trained. The negative correlation between difficulty and frequency indicated that very difficult behaviors occur less frequently than easier ones. The highest correlation was found between criticality and importance to train, suggesting that subjects feel that training is needed for these critical skills.

Table 1. Correlations Between Criticality, Difficulty Frequency, Importance to Train Ratings.

	Criticality	Difficulty	Frequency	Training
Criticality	1.00	.25*	.36*	.88*
Difficulty	----	1.00	-.37*	.27*
Frequency	----	----	1.00	.52*
Training	----	----	----	1.00

Note: $n = 134$, $* p < .01$

Results of this data collection effort indicated that pilots consider aircrew coordination behaviors to be important to their jobs. Overall, the aircrew coordination behaviors were rated as critical, frequently occurring, and important to train. The lower ratings on difficulty indicated that pilots do not consider aircrew coordination behaviors to be difficult to perform, but that they should be trained nonetheless.

Summary

Now that the importance and relevance of aircrew coordination behaviors have been established for this aviation community, they are being used as examples of the skills needed for the development and evaluation of a skill-based aircrew coordination training program. Based on independent classification by job experts these behaviors have been arranged under seven dimensions: Mission Analysis, Assertiveness, Adaptability/ Flexibility, Situational Awareness, Decision Making, Leadership, and Communication.

It is not known if these aircrew coordination behaviors are specific to the rotary wing community they were developed for, or if a fixed wing community could effectively employ the same behaviors. It is possible that the skills needed for effective aircrew coordination are very similar in a variety of aviation communities, and if so, a standardized aircrew coordination training program could be used by these various communities with minor modifications.

References

Carroll, J. E., & Taggart, W. R. (1987) Cockpit resource management: A tool for improved flight safety (UAL CRM Training). Cockpit Resource Management Training, (Tech. Report No. NASA CP-2455). Moffett Field, CA: NASA Ames Research Center.

Foushee, H. C. (1987). Dyads and triads at 35,000 feet: Factors affecting group process and aircrew performance. Cockpit Resource Management Training, (Tech. Report No. NASA CP-2455). Moffett Field, CA: NASA Ames Research Center.

Ginnett, R. C. (1988). Cockpit crew effectiveness from the inside out: A micro-analysis leading to macro-considerations. Proceedings of the Eleventh Psychology in the Department of Defense Symposium.

Jensen, R. S. (1987). Cockpit Resource Management Training, Final Report, Contract No. DTFA01-80-C-10080. Arlington, VA: Systems Control Technology, Inc.

Lauber, J. K. (1987) Cockpit resource management: background studies and rationale. Cockpit Resource Management Training, (Tech. Report No. NASA CP-2455). Moffett Field, CA: NASA Ames Research Center.

Morgan, B. B. Jr., Glickman, A. S., Woodward, E. A., Blaiwes, A. S., & Salas, E. (1986). Measurement of Team Behaviors in a Navy Environment (Tech. Report No. NTSC TR-86-014). Norfolk, VA: Old Dominion University, Center for Applied Psychological Studies.

Oser, R., McCallum, G. E., Salas, E. & Morgan, B. B. Jr. (1989). Toward a Definition of Teamwork: An Analysis of Critical Behaviors. (Technical Report No. NTSC 89-004). Orlando, Florida. Naval Training Systems Center.

Systematically Organized Workshops to Develop Computer Literacy Among Air Force Personnel

Russell N. Cassel, Colonel, A.F. Retired
CASSEL PSYCH CENTER
Heart Disease Risk Factor Clinic,
1362 Santa Cruz Court,
Chula Vista, California 92010

Abstract

Computer literacy is essential for effective functioning in the newly emerging age of communication and information. It involves more than being able to operate the computer, and must include some basic knowledge of the computer structure and technical functioning. A five module workshop is proposed for developing such computer literacy among select Air Force personnel, together with a parallel form of test to assess strengths and weaknesses both before and after the proposed workshops.

Peter Drucker, the Father of Management, maintains that today we live in a social and technological revolution, and that computer literacy is fast becoming our capital base (Drucker, 1989). In the newly emerging world of the 1990´s communication and information will be the watchwords, and computer literacy has become a universal need. Because of the great diversity of our present workforce and school curriculums, computer literacy too often receives little or no attention, and must be dealt with by the Air Force as a first order of business, if we are to meet our goals for national security.

The CPU Controversy

At the heart of computer literacy lies the "Computer Procssing Unit" (CPU), and which, of course is that part which processes the data. In this sense, then, it is the computer. For the great mass of microcomputers that have emerged on the scene over the past decade, this clearly involves two principal CPU´s: (Borton and Rossett, 1989):

1. The PC 88 series (80-88 an 8 bit processor, 80-286 a 16 bit processor, 80-386 a 32 bit processor, and 80-486 a 32 bit processor) made by the Intel Corporation.
2. The McIntosh series (6800 an 8 bit processor, 68010 a 16 bit processor, 68020 a 32 bit processor, and 68030 a 32 bit processor) made by the Motorola Corporation.

Intel has promised transparency between and among the later developments of the evolution in the CPU´s; while Motorola has not promised transparency. This, of course, means that software and programs purchased and running under one series of the CPU will continue to run under later developments, even though maybe slower and much more clumsily; while Motorola makes no such promise. Therefore, individuals who purchase the PC like machine can expect to continue to use there programs with the newer emerging CPU´s; while Motorola CPU´s would be faced with the threat of starting anew with each later revision of the CPU. There is much to be said for the advantages and disadvantages of each of these positions. This dialogue would make a computer literacy module worthy of concentrated attention, but remains a small portion of what constitutes computer literacy today.

The Functional Controversy

There is a group of individuals that maintain that computer literacy must be concerned primarily with how to run and use the computer, and not deal with the structure and functions of the computer itself. This, to be sure, could make a rather complex learning module, but which would lack depth concerning how the actual processing takes place. It might, of course, include an explanation of binary code, and how two digits lie at the basis of all computer functioning, and including the use of registers to manipulate data, but would not necessarily deal with other subtleties related to computer structure.

Proposed Computer Literacy Workshops

A series of five computer literacy workshop is proposed to foster computer literacy for appropriate Air Force personnel. Each of the separate workshops is planned to cover an essential area of development and evolution, and to cover an evening and the following day of learning. This, to be sure, is completely independent of hands-on experience, and where the needs will vary greatly depending on the workstations of the personnel involved. For some the experience might well be in the word processing area, for others it might be data-base management, for still others the use of higher level languages, etc.(Benderson, 1981):

WORKSHOP I -- Computer Evolution and Historical Development.
WORKSHOP II -- Computer Structure, Jobs, and Languages.
WORKSHOP III -- Data Processing and Statistical Analysis.
WORKSHOP IV -- Data Management Systems and Memory Retrieval.
WORKSHOP V -- Word Processing and Communication.

WORKSHOP I -- Computer Evolution and Historical Development

This workshop is intended to introduce the evolutionary development of the computer and historical events and things related thereto. This module includes a summary of source information in three areas: (1) people, (2) breakthroughs, and (3) vocabulary. The outline for this module is as follows:

1. Modern technology and computer.
2. Critical people in computer evolution.
3. Critical "breakthroughs" in computer evolution.
4. Specialized vocabulary related to computer evolution.
5. Impact of society by computer.
6. Nature and function of computer literacy.
7. Matching exam for critical people.
8. Matching exam for computer breakthroughs.
9. Matching exam for computer vocabulary.
10. Computerized Computer Literacy Test (COMLTC1).

WORKSHOP II -- Computer Structure, Jobs, and Languages

This module follows WORKSHOP I logically, and seek to introduce the functioning elements of the computer, the jobs that are involved in computer applications, and languages utilized. A summary of the contents is as follows (Duncan, 1989):

1. Computer generations.
2. Computer sizes.
3. Computer system.
4. Computer structure.
5. Computer secondary memory.
6. Disk storage devices.
7. Line printers.
8. System software.
9. Computer careers.

WORKSHOP III -- Data Processing and Statistical Analysis

This module seeks to cover the principal broad specialized areas for use and applications of the computer, and a summary of the contents is as follows:

1. Statistical analysis of data.
2. Architecture and engineering.
3. Mechanical simulation.
4. Simulation in the behavioral sciences.
5. Guidance systems.
6. Robot utilization.
7. Astronomy and space use.
8. Forecasting and predictions.
9. High tech toys.
10. Computerized music.
11. Computerized art.
12. Medical diagnoses, and monitoring.
13. Pattern making.
14. Artificial intelligence.

WORKSHOP IV -- Data Management Systems and Memory Retrieval

This module focuses largely on school use and learning paradigms, and a summary is as follows (Haehl, 1989; Cutler and Truss, 1989; and Stone, 1989):

1. Learning paradigms.
2. ERIC System.

3. Police systems.
4. Electronic spreadsheets.
5. Advanced accounting systems.
6. Vocational guidance systems.
7. Library management systems.
8. Computer assist instruction.
9. Computer managed instruction.
10. Case study analysis.

WORKSHOP V -- Word Processing and Communication

This workshop seeks to focus on the usual office support systems that have emerged through the use of the computer, and a summary of the contents is as follows (Mendelson, 1989; Van Name and Catchings, 1989; and Robinson, et.al., 1989):

1. Word processing.
2. Anagrams for crossword puzzles, writing poetry, etc.
3. Information guide.
4. Administrative applications.
5. Media and publications.
6. Electronic banking and stockbrokerage.
7. Electronic shopping.
8. Audio and video support units.
9. Language translators.
10. Telecommunication.
11. Security and privacy, and criminals.
12. Computer Literacy Test (COMLTC2)

Computer Literacy Tests

Two parallel forms of a computer literacy test have been compiled and which have been computerized. They are intended for use before and after the five workshops to determine information gained in relation to computer literacy. Each of the tests have six part scores to depict strengths and weaknesses in relation to computer literacy of students involved. Items are distributed as follows:

	COMLTC1	COMLTC2
PART I - Computer Utilization:		
1. General information	5	5
2. Higher level languages	5	5
3. I/O devices	5	5
4. Computer operations	5	5
5. Computer hardware	5	5
TOTAL	25	25
PART II - Word Processing:		
1. General information	5	5
2. Editing commands	5	5
3. File and block commands	5	5
4. Dictionary and spelling	5	5
5. No file commands	5	5
TOTAL	25	25
PART III - Data Base Management:		
1. General information	5	5
2. Changing data base files	5	5

3. Output of database	5	5
4. Manipulating database files	5	5
5. Data banks	5	5
TOTAL	25	25
PART IV - Trend/Simulation and Spread Sheet:		
1. Trend and analysis	5	5
2. Simulation	5	5
3. Spread sheet	4	4
TOTAL	14	14
PART V - Telecommunication and Guidance:		
1. General information	4	4
2. Computer conferences	5	5
3. Guidance systems	5	5
TOTAL	14	14
PART VI - Graphics/Speech/Motion:		
1. General information	4	5
2. Windows	5	5
3. Motion, speech, and color	5	4
TOTAL	14	14
GRAND TOTAL	117	117

References

1. Benderson, A. (1981). Computer Literacy. Princeton, New Jersey: Educational Testing Service.
2. Borton, W., and Rossett, A. (1989). Educational software and published reviews:Congruence of teacher developer and evaluator perceptions. Education. 109, (4), 434-444.
3. Cutler, R.B., and Truss, C.V. (1989). Computer aided Instruction as a reading motivator. Reading. 26, (2),103-109.
4. Drucker, P.F. (1989). How schools must change. Psychology Today. 19, May.
5. Duncan, R. (1989). Power programming: Use of algorithms. PC Magazine.8, (21), 337-338.
6. Haehl, S. (1989). ERIC/RCS. Journal of Reading. 33, (2), 146-147.
7. Mendelson, E. (1989). Word processing. PC Magazine. 8, (20), 97-100.
8. Robinson, M.A., Gilley, W.F., and Uhlig, G.E. (1988). The effects of guided discovery LOGO on SAT performance of first grade students. Education. 109, (2), 226-230.
9. Stone, M.D. (1989). Spreadsheet analysis. PC Magazine. 8, (20), 197-198.
10. Van Name, M.L., and Catchings, B. (1989). Networks: The mailman cometh.BYTE. 14, (9), 143-146.

The Impact of Human Factors on the Acceptance of Microcomputer-Based Modeling Tools

Capt David P. Martin, M.S.
Lt Col Tim O. Peterson, Ph.D.
Air Force Logistics Management Center

Abstract

Microcomputer-based modeling tools have the potential to provide Air Force program managers with significantly better cost/benefit analysis capabilities than past paper and pencil methods of performing these analyses. However, in an effort to model the real world accurately, model builders often ignore the human factors which determine if the models will be accepted by the user. This paper highlights the human factor problems observed in several of the Air Force microcomputer-based modeling tools and recommends specific areas to which model builders should devote more attention if these models are to gain wider acceptance.

Microcomputer-based (micro-based) modeling tools developed in the 1980s have the potential to provide Air Force program managers with revolutionary mechanisms for conducting more rigorous and extensive economic analyses of various logistics support alternatives for weapon system acquisitions. In addition, these modeling tools can reduce the time necessary to complete exhaustive cost/benefit analyses. Recent microcomputer enhancements allow for the development of powerful micro-based models. A decade ago, these models were only available on large mainframe computer systems.

The use of micro-based models has dramatically increased in the past five years (Jeffery, 1989). Unfortunately, the enormous potential benefits have not been realized. Many of the obstacles that prevent these models from gaining wider acceptance are the direct result of the model builders' dismissal of the importance of the human factor in the modeling equation. Carey (1988) argues that for complex tasks, such as economic analysis, the user is an integral part of the computer system. If the system is going to be accepted by the user, then it must enhance the problem solving environment, not degrade it. So, model builders must design these micro-based modeling tools to fit the user, instead of forcing the user to fit the modeling tool (Woodson, 1981). This can only be done by making the human element an essential factor in the model development equation.

The Limitations on Model Acceptance: Human Factors

Human factors is the scientific study of the interaction between people, machines, and their work environment (Beard & Peterson, 1988). In this case, it is the psychology of information systems. When findings from human factor studies are applied to the design of information systems, the friction between the user and computer system is greatly reduced (Scheiderman, 1987). However, when these findings are ignored or blatantly dismissed, dysfunctional behaviors arise as described by Peterson and Peterson (1988).

As part of an earlier study (Martin, 1989) seven micro-based modeling tools were evaluated to determine their feasibility in enhancing cost/benefit analyses in Air Force program offices. During this earlier study, a number of human factor issues surfaced. In this manuscript, the authors present several examples of the impact of poor human factor design which hindered model acceptance among program managers at major Air Force program offices. These examples fall into the following categories: 1) training difficulties, 2) data collection difficulties, and 3) data entry difficulties.

Training Difficulties

Effective use of micro-based modeling tools goes beyond an understanding of how to merely "push the proper buttons." In addition to fundamental microcomputer skills, the model user must also possess a cognitive (mental) model of the process which has been automated. Without this cognitive model, the user becomes nothing more than a parrot, a "user that executes commands when told to without any understanding" (Trumbly & Arnett, 1989: 4).

Beyond a basic cognitive model and some basic computer skills, it is also necessary for the model user to have some rudimentary knowledge of the relationship between the input and output parameters. This knowledge allows the user to have an essential understanding of how various model inputs can be combined to produce specific model-generated outputs. This doesn't mean that the user must understand every mathematical formula and their relationships within the model. Ideally, the proper validation and verification steps identified by Sargent (1987) and Hallam et. al. (1987) have been completed prior to the release of the micro-based modeling tool. Nevertheless, proper model use does require the user to peer inside the model structure to see how it operates. Without this mental representation of the process, it would be almost impossible for the user to determine if his/her input data had combined to produce a logical output or nonsensical output.

As the acquisition of microcomputers throughout the Air Force continues, ensuring that microcomputer users obtain a sufficient level of basic computer knowledge can be a difficult task. Nevertheless, there is evidence that the Air Force is developing a microcomputer-literate work force with the integration of microcomputers into entry-level technical training courses and into academic programs such as the Air Force Academy and the Air Force Institute of Technology.

However, in developing expertise for micro-based modeling use within the Air Force program management environment, our observations indicated that training for program managers on how to use these micro-based tools is often poorly done or non-existent. For example, one program office revealed that not one of the program managers responsible for performing cost/benefit analysis was using a micro-based modeling tool to assist with this task even though the micro-based tools were available. In addition, interviews conducted with other Air Force program management offices indicated similar nonuse of these micro-based modeling tools.

The lack of available training was cited as one of the primary reasons for not using these modeling tools. As one program office division chief put it, "My program managers are responsible for the oversight of between ten and fifteen active projects. They don't have time to sit down and read through a set of complex user manuals to get 'up to speed' on how to use one of these models." All too often, the development, verification, and validation of the micro-based modeling tool is the sole focus of the model builder with no regard for human factor issues. Many times, the only way for a user to learn how to use these modeling tools is to read the manual or to experiment with the modeling tool. Needless to say, given the day-to-day pressures of the Air Force acquisition environment, many program managers are hard pressed to find any extra time to devote to self-instruction in micro-based modeling tools. As a result, many times the models go unused by the people they were intended to assist.

Data Collection Difficulties

Micro-based modeling tools vary greatly in their complexity. At one end of the complexity continuum are user-developed models tailored for a specific need. Often these user-developed models are developed using commercially available spreadsheets. On the other end of the continuum are composite micro-based models. These composite models integrate two or more existing, but smaller, models to facilitate the capture of multiple measures such as reliability, maintainability, availability, or life-cycle costs.

While the more complex models provide the model users with more extensive analyses, they also require the user to enter a significantly larger amount of data. In an effort to develop a more complete model of the world, many of the complex model builders have left the data collection effort entirely up to the user. Past research (O'Reilly, 1982) shows that when the user is tasked with data collection, the primary criteria used in selecting the data is accessibility of the data.

With one notable exception, all of the micro-based models observed were developed independently of existing Air Force data systems. All of the model builders assumed that the user would know where to get the data needed to operate the particular model. This assumption usually results in users having to locate and thoroughly familiarize themselves with several different Air Force data systems. Without such familiarization, the successful use of the micro-based model is not possible.

In the day-to-day operational environment of most Air Force program management offices, such a data system familiarization process is simply not possible. As a result, after program managers (model users) struggle for awhile with the tremendous amount of data required for many of the more complex models, most abandon the use of the more complex modeling tools altogether. Often this leads to a souring by the program manager on the whole micro-based modeling experience.

Data Entry Difficulties

Another human factor which receives minimal attention from model builders is data entry. Scheiderman (1987) points out that data entry can be a source of frustration and potential errors. For this reason, data entry requirements for micro-based modeling tools should be given considerable attention. It can be a very frustrating experience to have the model user spend several days or even weeks collecting the data needed to operate the modeling tool, only to encounter additional problems in trying to set up the data for the modeling tool.

Smith and Mosier (1984) suggest that a guiding principle for data entry is to minimize the input action of the user. By doing this, greater user productivity is realized and also less chance of data entry errors. In one of the modeling tools observed, over 2500 inputs had to be made before the model could be executed. Smith and Mosier (1984) also argue that redundant data entry should be avoided at all cost. In the example cited above, the model required the rekeying of the same ten data elements 250 times to make the model executable. In an effort to measure the magnitude of this data entry task, one of the authors spent over five hours just entering the data required to run this model. Is it any wonder that micro-based models of this type are placed on the shelf and go unused?

Another data entry difficulty encountered was with the data entry format. In fact, the data entry format of two of the models observed was the primary reason they were rejected for an earlier feasibility study (Martin, 1989). These two models required the construction of separate American Standard Code for Information Interchange (ASCII) text files, complete with data entries in all the correct places within the file. This data entry format procedure makes the faulty assumption that the model users will understand how to perform this esoteric task, and that once they understand how to perform the task, they have the patience to accomplish it. Since most microcomputer users are application users and not power users, they will quickly abandon the use of these models in favor of those that require less of a data entry burden, even if the model does not provide as accurate a representation of the real world.

Implications and Summary

While micro-based modeling tools exist which give consideration to one or more of the human factors mentioned, it is clear that additional emphasis must be placed on human factors before these models will become widely accepted for use in an Air Force program management environment. Many of these models are quite complex, and in the absence of any formal training program, many of the model users do not have the extra time required for self-instruction. It is possible that embedded training systems or computer-based training systems which simulate the modeling process could be excellent tutors for inexperienced users.

In addition, while many of these models accurately reflect reality, their concentration on details often places unrealistic data collection burdens on program managers. Ideally, a balance needs to be struck between model accuracy and data collection requirements. The best micro-based modeling tool observed during the feasibility study (Statistically Improved Life Cycle Cost Model) not only strikes this balance, but also guides users in their data collection effort. For example, this modeling tool provides the user with the data collection requirements, as well as the name and phone number of the agency having access to the required data.

Furthermore, model users are often burdened with trying to determine the proper way to enter data into the model. Only two of the seven models examined provided the user with an effective data entry screen that included adequate on-line help for each data entry and reduced excessive repetitive data entries.

In summary, microcomputer-based modeling tools are still in the developmental stage. The primary thrust of model builders to this point has been to deliver models that accurately reflect reality. While this is important, the price for this realism has been the neglect of the model user in the areas of training, data collection, and data entry. Unless more emphasis is placed on these important human factor issues in future, it will be some time before these micro-based modeling tools make inroads into the Air Force program manager environment.

References

Beard, J.W. & Peterson, T.O. (1988). A taxonomy for the study of human factors in management information systems (MIS). In Jane M. Carey's (Ed.), Human factors in management information systems. Norwood, NJ: Ablex Publishing Corporation.

Bernard, M. (July 21, 1989). Personal Interview. Chief, Tactical Systems Division, PRAM Program Office, Wright-Patterson AFB, OH.

Carey, J.M. (Ed.) (1988). Human factors in management information systems. Norwood, NJ: Ablex Publishing Corporation.

Hallam, D.D. and others. (August 1987). The validation of simulation models. Data Management, 25, 83-87.

Jeffery, A. (May 28, 1989). Personal Interview. Manager, SIDAC Concept Exploration Team Project, Dynamics Research Corporation, Fairborn, OH.

Martin, D.P. (1989). Using Microcomputer-Based Logistics Models to Enhance Supportability Assessment for the USAF Productivity, Reliability, Availability, and Maintainability (PRAM) Program Office: A Tailored Approach. (MS Thesis, AFIT/GLM/LSM/89S-40, Air Force Institute of Technology, Wright-Patterson AFB, OH).

O'Reilly, C.A. (1982). Variations in decision makers' use of information sources: The impact of quality and accessibility of information. Academy of Management Journal, 25, 756-771.

Peterson, C.M. & Peterson, T.O. (1988). The darkside of office automation: How people resist the introduction of office automation technology. In Jane M. Carey's (Ed.), Human factors in management information systems. Norwood, NJ: Ablex Publishing Corporation.

Sargent, Robert G. (1987). An overview of verification and validation of simulation models. Proceedings of the 1987 Winter Simulation Conference, 33-39. Association for Computing Machinery, New York, NY.

Scheiderman, B. (1987) Designing the user interface: Strategies for effective human-computer interaction. Reading, MA: Addison-Wesley Publishing Company.

Smith, S.L. & Mosier, J.N. (1984). Design guidelines for the user interface for computer-based information systems. Bedford, MA: The MITRE Corporation.

Trumbly, J. & Arnett, K. (1989). The performance relationship model: A framework for human factors research. Proceedings of the Second Symposium on Human Factors in Management Information Systems. Sacramento, CA.

Woodson, W.E. (1981). Human factors design handbook: Information and guidelines for the design of systems, facilities, equipment, and products for human use. New York, NY: McGraw-Hill.

The Plebe Interface

Kay Gladwell Schulze, Ph.D.
LCDR William G. Borries, USN
LCDR Dominick W. Visco, USN

Computer Science Department
United States Naval Academy

Abstract

Recent research has shown that an interface for software applications is more effective when it considers the communication needs and memory load of the user and provides a consistent model of the system. This paper describes a mouse driven menu interface for a microcomputer running the MS-DOS operating system. The interface was developed and is currently being used by the class of 93 at the United States Naval Academy.

The advent of mandatory, student owned microcomputers in the academic environment has created the need to insure that students indeed use them. This problem has not arisen in the past mainly because students generally had only limited exposure to computers in high school and few computer illiterate students owned their own computer. However, with more and better educational software available, computer usage has filtrated down into the elementary schools. As a result more and more college freshmen have had exposure to the capabilities of the computer and several universities now encourage or require them to purchase their own microcomputer.

Since the computer is now an integral part of the United States Navy, the United States Naval Academy, commencing with the class of 90, has required each plebe to purchase a personal microcomputer. It was soon discovered that ownership did not necessarily ensure usage. Since their inception, the overall curriculum of the Academy has been strengthened to incorporate the computer in all aspects of the midshipmen's academic and professional endeavors. Even though many of the midshipmen have had prior exposure to computers, their experience with MS-DOS has generally been limited. As a result, it became necessary to rapidly and effectively transform approximately 1500 plebes into computer literate users. To this end, the Academy began developing various interfaces for the MS-DOS environment.

History

The initial attempt to acclimate the midshipmen to their computers was to provide a single USNA Menu. Since it was a yard wide problem, this interface was developed by Computer Services. It presented the user with a list of options, driven by function keys, that launched each of the major software applications in use at the Academy. However, once the application was launched, the midshipmen found themselves in the MS-DOS command line environment which also required them to remember the correct path name of all the files.

A task can be completed more easily when the amount of information necessary to perform the task is kept at a minimum. This information is stored in a person's long term memory and must be transferred to his short term memory. In order to reduce this load on the midshipmen's working memory, the Computer Science Department introduced a software application called Directory Scanner™ [1]. This application is a graphical display of the hierarchical file system which by using function keys, allows the user to traverse the hierarchy of directories. Once he located the desired directory, pressing the *Enter* key displayed the files contained in it. He could then select the desired file by using the arrow keys. But once again, the midshipmen were dependent on function keys and the command line to accomplish any useful work.

Since long term memory encodes material that is meaningful, students are normally not as proficient at remembering facts as they are at remembering rules. The MS-DOS operating system is often associated with the ability to remember facts whereas a mouse is more frequently associated with rules (Thomson, 1984). Based on this, it became evident that an interface, more dependent on the use of a mouse, needed to be developed.

One of the obstacles to a mouse driven interface was the lack of memory and disk space available on the plebe machine. However, this difficulty disappeared with the class of 93, when each plebe was required to purchase a Zenith 286 LP microcomputer with one megabyte of RAM, a 20 megabyte hard disk drive, a 1.4 megabyte single floppy disk drive, a VGA color monitor and a Logitech mouse. The Computer Science Department and Computer Services undertook to develop a mouse driven interface to facilitate the plebes' acclimation to the machine(Geschke, Borries, Visco and Rowe, 1989).

---

[1] Directory Scanner™ v 3.3 was developed by Nat Martino.

The Plebe Interface

After a review of the literature of current interface designs, we decided to develop and implement a closed design that relies totally on the mouse. The metaphor for a closed system was first presented by Laurel (1986). Such a system behaves deterministically and no important aspects of its behavior are omitted. In such a system, the user should be able to form a mental model of the system that is complete and obvious. Hence the actions the user learns to perform the first time he uses the system will cause him to form a model of the system that will help him in using the system in the future. Any new actions that he encounters will be explainable in terms of the model.

The plebe interface is developed on top of Directory Scanner™ and allows for traversal through both the directory hierarchy and the corresponding files using the mouse. However, to ensure acceptance among users who were already adept with the MS-DOS environment, the entire interface can be manipulated using function keys, etc. This fact, however, is not stressed to the naive user.

The Logitech mouse consists of three buttons. The left button is used for *Escape*, the middle one will bring up the *Help Menu* and the right one designates *Enter*. Four chords are available by combining buttons. Pressing the left and middle buttons will toggle the directory section of Directory Scanner™. The midshipman then moves the mouse to the desired directory and selects it using the right mouse button. This action causes the files in that directory to be displayed. He can then move the mouse to the desired file and select or tag it by pressing the middle and right mouse buttons.

The interface maintains a Main and a File Command Menu for Directory Scanner™ and for the other software applications in use at the Academy. As software applications are added or changed, they can be accessed by the plebe interface once menus are written for them and the Directory Scanner™'s Main Menu is modified. The Main Menu is brought up by pressing the left and right mouse buttons and contains options for the major tasks pertinent to each application. For example, the Main Menu at the Directory Scanner™ level contains options that launch applications such as Quicksoft's PC-Write™, Borland's Turbo Pascal™ and Quattro™, ProComm™ and an in house Math Plotting Package. In launching an application, the interface removes Directory Scanner™ from memory, loads the menus for the application and then executes the application. Upon exiting, it reloads the Directory Scanner™ menus and Directory Scanner™. The Main Menu also provides options for formatting data and system disks and for shutting down the system.

The Main Menu for each of the applications is comparable. For example, in the PC-Write™ application it contains options for cutting and pasting, searching and replacing, traversal of the text and formatting the text. In the Turbo Pascal™ application, options are provided for running or compiling a program, several debugging techniques and saving a file. The ProComm™ menu provides for a system break, a redisplay of the last screen and an exit to and from MS-DOS. The Math Plotting Package menu provides options for graphing integrals, slopes, contour lines, vectors and determining roots to equations. The builtin menus for Quattro™ were sophisticated enough that the plebe interface simply defaults to them. Each of the Main Menus provide an *Exit* option that returns the student to Directory Scanner™ and an option that invokes the File Command Menu.

The File Command Menu can be brought up either from the Main Menu or by pressing all three mouse buttons. The options in the File Command Menu at the Directory Scanner™ level are the typical MS-DOS commands that allow for files to be listed, copied, moved and deleted. In addition, it provides options for modifying directories and changing disk drives. Several of the commands will bring up a second menu. For example, when the *Copy File* option is selected, another menu providing various destinations will appear.

The File Command Menu for PC-Write™ provides options for inserting and printing files and various save and exit commands. In Turbo Pascal™, the menu provides various editing commands while in ProComm™, it provides for opening and closing log files. The Command File Menu for the Math Plotting Program is embedded in each of the Main Menu options and is still driven by function keys. The first option in each File Command Menu is *Exit* which returns the user to the application.

Summary and Conclusions

Two Saturdays prior to the commencement of classes, the plebes were issued their computers, the software applications and a Quick Start manual (Welcher, 1989). The first Saturday, under the supervision of a technical advisor from Computer Services or the Computer Science department, the plebes set up their machines and started on the Quick Start manual. This gave them an introduction to the use of the mouse with Directory Scanner™ and allowed them to use ProComm™ to connect to the Naval Academy Data Network. The second Saturday, under the supervision of Computer Services and the Chemistry department, they were given a Quattro™ Quick Start manual (Pearson, 1989) and an introduction to the use of the software.

During the first week of classes, the Naval Leadership and English departments required each plebe to submit a paper that had been written using the PC-Write™ word processing application., Half of the plebe class also started the required Fundamentals of Computing course which immediately reviewed all of the software applications and then immersed them into pascal. By the end of the second week of classes, most plebes were able to use their computers in a productive manner.

The plebe interface was developed to reduce the memory overload that occurs when a midshipman is first introduced to a computer operating system and various software applications. Although no empirical studies have yet been concluded that verify the effectiveness of the interface, all indications are that the class of 93 made the transition from novice to computer literate in a more rapid and effective manner than any of the prior classes. Further studies are indicated to determine whether the interface facilitates the midshipmen's continual usage of their computer.

References

Geschke, M. J., Borries, W. G., Visco, D. W. and Rowe, W. R. (1989). Computer science disk scanner interface and documentation. Unpublished manuscript, United States Naval Academy, Computer Science Department, Annapolis.

Laurel, B. L. (1986). Interface as mimesis In D. A. Norman & S. W. Draper (Eds.), User centered system design (pp. 67-85). New Jersey: Lawrence Erlbaum.

Pearson, W. H. (1989). Quattro™ quick start. Unpublished manuscript, United States Naval Academy, Chemistry Department, Annapolis.

Thomson, N. (1984). Human memory: Different stores with different characteristics In A. Monk (Ed.), Fundamentals of human-computer interaction (pp. 49-56). New York: Academic Press.

Welcher, P. J. (1989). Quick start. Unpublished manuscript, United States Naval Academy, Mathematics Department, Annapolis.

Installing the Integrated Maintenance Information System Into a Portable Flightline Aid

Lt Cher E. Wynkoop
Air Force Human Resources Laboratory
Logistics and Human Factors Division
Wright-Patterson AFB, OH

Abstract

The purpose of this paper is to describe the developmental process of designing the physical aspects of a prototype interface device to fulfill the Integrated Maintenance Information System concept. The paper describes a history of the requirements analysis, details software toolkit design, discusses keyboard options and arrangements, and outlines user needs for the physical design of a Portable Maintenance Aid (PMA) for use by maintenance technicians.

For over twelve years the Air Force Human Resources Laboratory at Wright-Patterson Air Force Base has been conducting research to develop methods of presenting technical information on a portable computer to support its Integrated Maintenance Information System (IMIS) development program. The IMIS concept is one that will provide an aircraft maintenance technician with a single computer interface to the information and data bases that he needs to access and interrogate to accomplish his work. The technician will use a hand-held portable computer that interfaces directly with existing maintenance data bases and the multiplex data bus of the aircraft. It will perform built-in tests, read and analyze fault data, provide diagnostic advice, and present automated technical procedures. IMIS will directly impact sortie generation in a very positive manner by improving the performance of the maintenance technician.

The Laboratory initiated its investigation of presentation requirements for automated technical information by developing a prototype computer system to support maintenance activities in the shop. The prototype was used to demonstrate the feasibility of the concept and to conduct research on specific human factors issues. The in-shop prototype was located in a sheltered area, so a full size screen for presentation and a keyboard for data entry were used. After successfully completing the in-shop tests, the lessons learned were carried over to the development of a prototype computer for flight line use. The flight line work environment induced severe size and weight restrictions. The prototype computer must be small and easily portable; thus, the screen size and number of keys must be limited. These restrictions impose many unique and difficult human computer interaction requirements. The Laboratory is designing an interface that will support various levels of computer expertise, varying levels of experience on specific maintenance tasks, and will be suitable for a variety of maintenance tasks. The final design of the PMA must incorporate the general needs of the user and the more specific human factors engineering issues that will lead to a user-friendly system. This paper describes the procedures used in designing a human computer interface that meets the needs of the aircraft maintenance technician. At this stage of development of the PMA, the initial layout is completed, but there are several issues concerning the presentation system and interaction techniques which are being investigated and adapted as current technology evolves.

Method

The Laboratory has taken several steps in developing a human-computer interface design for a portable maintenance aid. A functional analysis of the users' requirements was performed by human factors engineers (HFE). The HFEs established a fundamental set of system requirements from interviews with maintenance technicians, experience on previous projects, and the reviews of other automated systems. The HFEs determined that a graphical user interface (GUI) would provide the most effective means of meeting the technicians' needs and the Laboratory's design goals. After making the decision to use a GUI, the HFEs defined the fundamental software elements needed to build the software toolkit required to develop the GUI. A software toolkit is used to build the application software for the prototype.

The next phase in our prototype design is to develop maintenance scenarios and screens on a computer with rapid prototyping capabilities including hypertext links. Using these powerful linking mechanisms, we can visualize worst-case scenarios on the flight line and identify what resources technicians would need from the computer in each maintenance situation. By doing this, we can enhance our initial software toolkit to include a more detailed arrangement of interactive elements.

Our initial software toolkit consists of elements mainly for navigation, sequencing, and data entry. A few examples are SCROLLER, BUTTON, MENU, DIALOG BOX, and WINDOW. A SCROLLER provides the capability for viewing an illustration that is larger than the screen. A BUTTON is a mechanism that allows the viewer to choose a response. A MENU is a list of options the viewer might have in any one scenario. A DIALOG BOX is another mechanism that allows the viewer to choose one of several responses. The WINDOW is the basic frame inside which all activity occurs. The WINDOW has several components. WINDOWS have frames, an area in which to read and interact with data, title bar, and a scroller. The WINDOW can be altered in size, location, or content by the programmer as well as the technician. The user must navigate through, sequence, and input the data presented in the windows by moving the cursor to enter his commands. This cursor interaction is required due to a graphical user-interface and accomplished through various input devices such as a keyboard, mouse, touch pad, touch screen, track ball, and joystick.

An input device that is fast, accurate, easy to use, and appropriate for use in the maintenance environment is needed. The mouse was considered inappropriate for our PMA in a flight line environment because the PMA must be a single, self-contained unit. A mouse would be an extra device that the technician would have to carry, along with his tools. The joystick, the touch tablet, the touch screen, the track ball, and the keyboard are being evaluated specifically for the maintenance environment. Size is a critical issue in considering input alternatives.

Specifications, design considerations, and technology limitations dictated that the box be about five pounds in weight, with a screen size of 6" by 8". The computer must be suitable for carrying and operating with one hand. The joystick, according to most engineers in the branch, would be damaged in flight line handling since it would stand above the surface of the PMA and is very fragile. A touch screen is a possibility. However, there is concern that dirt and grease on the hands of the technicians will obscure the display. Also, there is a risk of damage to the display if the technicians should use their tools to "touch" the screen. A touch tablet and trackball could have the same degraded functionality due to conditions in a dirty environment and would add extra weight to the PMA. Although the keyboard entry method appears to be the most viable method of entering information into the PMA available at this time, we will research the aforementioned devices further.

Regardless of the choice of input devices, the maintenance technician should not be required to use two hands for data input. In fact, one of the goals behind this PMA design was to eliminate any need for the technician to have extensive typing skills. The PMA keyboard should have no alpha character keys, and it should allow the technician to operate the machine with one- or two- button presses.

Once we had established the keyboard as the most likely input device, it was necessary to determine how many keys would be needed, which keys would be dedicated, which would be software programmable, and how they would be arranged on the computer. We then had to define the function of those keys which were not directional. We first listed all the functions the computer has to perform based on the different types of information displayed and users' needs. Based on the knowledge of the maintenance staff and the results from several field tests and interviews with technicians, we identified critical functions for which dedicated keys would be provided. Dedicated keys would be reserved for those functions that the technician wanted to perform often and that had no branches or further menu items attached to them. For example, NEXT, BACK, and MENU BAR should be dedicated keys because when the technician wants to go to the next frame, he/she should be able to do so with one singular action on the keyboard. With one press of NEXT, the technician could view the next resource available. For more infrequent actions like changing the font or for actions that could have many different alternative choices, (e.g., calling for HELP, the technician can obtain this information with a combination of numerical keys or selecting options from a menu. We also included six function keys on the keyboard to act as dynamic options that the technician would want to select with only one keystroke. Dynamic options are functions which are not always available. Their availability is specified in the data itself when they are available. They are identified in a box on the window. A few of these functions are SCROLL, ZOOM, CROSS REFERENCES, BOOKMARK, YES/NO, and HELP.

In conjunction with the function key decisions, we had to determine the key arrangement on the computer. The basic objective was to minimize the size of the computer. Should the computer be made as physically small as technology will allow, without regard to ease of use? Or, should it be made somewhat larger to allow more convenient placement of keys? There were arguments for both sides. The main problem the technicians had pointed out during evaluations of previous prototypes was that they were too large and awkward to handle in small areas. Basically the design engineers and the human factors engineers had to compromise. The box would have to be as small as possible, but the number keys had to be arranged in a logical order. The direction keys had to be kept in a usable arrangement, and the keyboard had to be somewhat similar to what people are accustomed to viewing and to using (e.g., typewriter, tabletop computers, etc.). With a growing body of technology and experience, our team of engineers will be able to decrease the size of the interface without sacrificing comfort and ease of use to the technician.

With the needed functions developed, we began to arrange the physical design of the PMA which included how the computer was held, its weight, its maneuverability, comfort of the user, and general human factors design considerations. We began this task simultaneously with all the others by taking a survey of the technicians at Homestead AFB, Moody AFB, Springfield Guard Unit, and the 4950th Test Wing at Wright-Patterson AFB. The engineers in the Laboratory also added Human Factors issues, but those issues of importance to our engineers were not related to those of the maintenance technicians. The Human Factors specialists were concerned with key placement (Fitt's Law), balance of the computer, order and arrangement of number keys, and input device selection. But the technicians did not respond to those concerns. Instead, they wanted ease of transportability; they wanted ruggedness; they wanted simplicity; they wanted a non-repair item; they did not want this portable device to be a hazard while moving around the aircraft. They wanted to be able to hold the computer with one hand, sit it on the ground, or hook it on a tripod-like device. In the final design of the PMA, our engineers will have to ensure the general interests of the technicians are met, along with the more specific yet vital concerns of the Human Factors specialists.

Conclusion

The development of the human-computer interface is an evolving process. The first test of the PMA will be in the Fall of 1990. The lessons learned from this test will be evaluated and utilized to advance the larger IMIS concept, which will give the maintenance technician a single interface with which to access and utilize the information and data bases he requires to accomplish his/her work.

TARGET SIZE AND LOCATION EFFECTS ON TOUCH PANEL PERFORMANCE: LIMITS ON INTERPRETATION AND INFERENCE

Dennis B. Beringer
New Mexico State University

Abstract

Accuracy of input using touch panel devices is affected by a number of variables, particularly those relating to the target of the touch. A screening experiment was conducted to further examine the effects of target position and size upon accuracy of the touch input. Results suggest that error for right-handed users is least near the resting position of the hand (lower right corner of display) and that response times were similarly affected. Accuracy was greater for targets demanding higher precision. It is recommended that for applications having established key input areas positions along the lower and right-hand borders of the control/display unit should be used to minimize activation time and error. Use of the lower border exclusively can accomodate users with either a right-hand or left-hand preference. Some comments are also provided on the limitations which bound the interpretation of results in several studies and inferences thus drawn.

There is currently an interest in defining guidelines for implementing touch-sensitive interfaces in Naval applications. The literature contains a variety of studies that have investigated variables affecting the accuracy with which individuals can use touch-input devices. The variables examined have included the operator's angle of regard (Beringer & Peterson, 1985; Hall, Cunningham, Roache and Cox, 1988; Beringer & Bowman, 1989), location of the target on the screen (Beringer; Hall, et al., 1988, Beringer and Bowman, 1989), technology type (Schulze & Snyder, 1983; Beringer, 1989), on-screen vs. off-screen input device (Whitfield, Ball and Bird, 1983) and other devices versus touch input (Karat, McDonald, and Anderson, 1986; Beringer, 1989). There have also been some examinations of the efficacy of training (feedback) in improving touch-input performance (Beringer & Peterson, 1985; Beringer, 1989) as well as studies of the effects of various types of gloves on operator performance (Beringer & Lee, 1988; Beringer, 1989).

All this data collected on numerous devices of varying technologies has produced sometimes complementary and sometimes conflicting results, or so it would appear. The apparent conflicts, however, may in many cases be a function of problems of interpretation or erroneous inference. Thus it may be useful, prior to reporting the results of yet another study, to examine some of the inherent limitations that may be overlooked in reporting and interpreting findings, not only of other researchers but often in our own work.

Limits of Interpretation and Inference

Error rates. Hall et al. (1988) suggested that no studies had previously provided activation error rates for touch panel operation. This is rather misleading in that an "error rate" is merely a function of the desired precision of activation and can be determined, as did Hall et al. after-the-fact in their Experiment 1, by applying accept/reject criteria to any collection of recorded responses. Data from earlier studies citing specific error data (x,y distance from target) could have been adapted to any application provided that some small central contact point was specified for each touch target (parity of visual targets). Thus the data that have been collected concerning input accuracy over the past decade are most likely applicable to a number of situations if some standard of interpretation can be adopted.

Mean effects and resolution. One must be extremely careful when citing data out of context. Hall et al. cite Beringer & Peterson (1985) as reporting a reduction in mean y-axis error, through training, of from -0.52mm to -0.20mm. Did this mean that the subjects could respond with this much accuracy? No. Resolution of the input device was only 3.2 mm. The data should really be interpretted as meaning that individuals were "on target" most of the time but occasionally were one touch cell off (about 1 in 15), and this is consistent with the data obtained in that study. In summary, error magnitudes that are appreciably less than input device resolution require special interpretation, most notably when resolution of the interface is comparatively coarse when compared with mean error (an order of magnitude difference in the Beringer & Peterson study).

An alternate way of examining this data would be to catagorize responses by touch units and to present a frequency distribution, either univariate or bivariate, representing the frequencies with which various magnitudes of error could be expected. This is particularly important for devices similar to the infrared panels having a resolution of only 1/8 inch. Inasmuch as error cannot be measured on a fine enough scale to be considered continuous, subsequent evaluation should treat the data in a categorical fashion. This appears to have been overlooked in a number of previous studies (e.g., Beringer & Peterson, 1985). Hall et al. came closest to this in their cumulative accuracy documentation, but the interface device they used was not of the type requiring data categorization (essentially continuous measurement for the purposes described). The same can be said for the data of Beringer & Bowman (1989).

It is apparent that one must use different devices to answer different questions. If one wishes to answer the question of how accurately one can place one's finger on a sloped surface, the appropriate measurement device may well be something on the order of the high-resolution thin-film resistive interface. If one wishes to determine how accurately a particular device can be used, one is then restricted to analysis and reporting in terms of that unit's resolution capabilities.

Interactions. One must also be cautious in an interpretational sense when reporting main effects for the variables of interest in these studies. The Hall studies reported a number of significant 2- and 3-way (and in one case 4-way) interactions but main effects were still discussed without being tempered by the presence of the interactions. The interactions were not fully explained, either. In this and other studies the result is that reliability of the main effects is clouded and one is often left to sort out when a main effect will go one direction and when it will go the other.

Range effects. There is also the notion of the "range effect" and the possibility that restrictions in the range of independent variables may produce "no-effect" situations. Previous studies had been concerned about the accuracy, in very-high-resolution terms, with which individuals could use a touch-input device. There was never any question among researchers as to the efficacy of the interface for menu selection using reasonably-sized key input areas. The Hall study did examine a limited range of targe size, limited to between 9.9 and 12.0 mm in the y axis. This range was not likely to elicit the potentially interesting effect, suggested by Beringer and Peterson (1985), that people may not touch accurately "...because their fingers get in the way." Thus one should be careful in interpretting no-effect results if the independent variable ranges are restricted.

Fatigue effects. It is interesting that Hall et al. found an orthogonal-regard mean error of -1.34 mm in y during the first exposure to a comparatively small target (+ character); that error grew to -2.29 mm in posttest (the 5th block of trials), replicating findings of earlier studies. It is also interesting that mean y error for the box-shaped targets varied from -1.5

mm to -1.71 mm with no significant effect of target size. These data fit a trend that is increasing in magnitude; all intermediate blocks show a higher magnitude of error (average of 1.60). Thus there appears to exist in these data the posited fatigue effect, observed previously, whereby error increases monotonically over trials.

Remaining Questions

It seems appropriate to examine a much wider range of target sizes, spanning the range from clearly obscured by the fingerpad to clearly visible beyond the bounds of the fingerpad. This is a more reasonable approach as a wider range of tasks are now represented and one can truly examine the relationship between desired precision and obtained accuracy (obtained accuracy should be a function of desired precision or, more succinctly, perceived precision requirements). This includes the case where view of target is lost just prior to contact.

Instructional set is also an issue here. Hall et al. instructed subjects to touch the centers of the targets presented. This can not be assumed to be standard operating procedure for all touch-panel users, particularly novices. In the most abstract sense, one could reason that a user must "touch the target". At the lowest level this can mean that some part of the fingerpad must contact some part of the target. The next higher level of interpretation would be that the contact point must be "within" the target, where a target has defined boundaries. At the highest level one sees the requirement as meaning that the contact point must be centered in or on the target. Each of these instructional sets should produce different precision levels in ascending order. Instructional set may not appreciably affect performance, however, if different instructions are given the same interpretation by the operator (i.e., "touch the target" being interpretted as "touch the center of the target").

The present studies were conducted to examine some of these unanswered questions, most noteably the questions of target size variation and instructional set. It was expected that variations in target size should produce variations in performance error, responses to smaller targets exhibiting smaller mean or variable error. It was also expected that variations in instruction to the operator should produce changes in obtained precision or variability with the above possible interpretational limitation.

Method

Design/variables

A modified central-composite design (CCD) was used to assess the effects of three variables (target location on the x axis, target location on the y axis, target size) at five levels each. The resulting economy in this screening design reduced data collection to 23 points from the 125 required for a single replication of a full-factorial design. This economy is not without cost, however, as points at the extremes of the experimental space are sampled once per replication with the majority of points being in the center of the variable space. Target size varied from 6.3 mm to 19 mm by 3.2mm increments. Target location was equally distributed about the center point of the display screen with a spacing of 160 display pixels (40 mm) between target centers. The CCD sampling strategy allowed nine unique combinations of x and y location to be sampled, four of these as part of a fractional factorial design. Categorical variables included in this study included gender, use of gloves, and presence/absence of visual feedback.

Apparatus

Both infrared (Carroll Touch) and resistive (Elographics) touch panels were used during data collection. The infrared device was mounted on a 14-inch flat tension-mask monitor while the resistive device was attached to a 13-inch curved-surface c.r.t. Each was driven/read by a dedicated microprocessor system. Given the number of variables addressed and space limitations, resistive-panel data will be given priority.

Subjects/Procedure

Twenty-four right-handed undergraduate students (12 male, 12 female) served as participants and used a thin-film resistive panel as the interface device. Each was seated before the appropriate display in an adjustable chair such that eye height was located on a line orthogonal to display center with the display within comfortable reach. Their task was to touch the square targets appearing on the display as quickly and accurately as possible. Instruction to this effect were presented as text on the display. Each participant performed four blocks of trials, two with the bare right hand and two with the gloved right hand. Trials were presented in four blocks of 23 conditions each, the order of these conditions being randomly selected. Visual feedback was provided on trials 2 and 4 indicating the registered input point relative to the intended target. The bare had was used during blocks 1 and 2, the glove being worn for blocks 3 and 4. Following data collection each individual was given an explanation of the purpose of the experiment and the intended application (AH-64B Apache helicopter).

Preliminary Results

Multiple regression for the x and y error measures (3 each) produced values of R^2 ranging from .13 (first x contact) to .20 (last y contact). Although significant predictors included gender, feedback, gloves, and some interesting second-order effects of size and relative target location in x for most of the dependent variables, only contact duration (R^2=.52) was well predicted by its combination of predictor variables. This was replicated in the second study for contact duration (R^2=.54) and log reaction time (R^2=.68), with group, subject, gender, and group by trial accounting for the vast majority of the variation.

Mean errors across all x,y locations ranged from +1mm to -1.5mm with typical standard deviations ranging from 2mm to 3.25mm (for both x and y error). Responding can thus be characterized as quite accurate. It appears that a square area approximately 20mm on a side would adequately account for the vast majority of variation found in this examination. This area is 33% smaller in each dimension than the tactual recognition field recommended by Hall et al. (1988) to accomplish the same accomodation of operator variability.

Reaction times (RT) across spatial locations on both the infrared and resistive panels followed similar patterns. Those targets closest to the top or left-hand side of the display produced the longest RT's and, for the infrared panel, greatest error. Contact durations were reasonably uniform across locations (mean = 130msec) excepting at the highest target on the screen where contacts were of shorter duration. RT's in the first study ranged from 504 msec to 652 msec. Although there were some statistically significant gender effects relative to accuracy and response time, the magnitude of these differences was so small as to make them nonsignificant in the practical sense.

The instructional variation result was interesting in that mean differences in x and y error by trial and group were of virtually no consequence (shift of 0.5mm at best). *Variation* in y error did decrease, however, for the experimental group overall (first point of contact: s = 12.2 to s = 8.6 pixels; from 3 to 2mm). Similar patterns were found for both RT and contact durations. RT increased significantly over trials for the experimental group (729, 893 and 1048 msec) as did variability, while RT decreased for the control group (611, 481, and 485 msec respectively). Contact durations decreased slightly over trials for the control group but did not change substantially for the experimental group. In summary, instructional emphasis on accuracy made subjects more deliberate and less variable in their responding but did not substantially reduce bias error (0.5 mm or so). This result combined with subjective reports from subjects suggests that individuals will, on average, interpret "touch the target" as "touch the center of the target". Instruction to do the latter, in most cases, is thus redundant.

Conclusions

The findings of this study suggest that the most frequently used "key" areas in menu selection should be placed near the lower or right-hand borders of the control/display area both to minimize activation time and to reduce magnitude of error. This is true for right-handed activation and should be mirrored in the x axis for left-handed activation. That variable error can be reduced by instruction was not surprising, but the improvement was so small as to suggest that most individuals are already performing as accurately as is possible for them. The apparent failure to account for much of the variation in x and y error is undoubtedly reflected in the fact that no extreme values of x and y position were used (near display edges) and the placement of most targets was within 4 cm of center screen (8 cm at worst possible case). This is, in fact, the good news, suggesting that one can expect some uniformity of precision across the display surface with this type of interface. It is also clear that the bias error found in previous studies of infrared devices is largely absent with the resistive panel, most likely as a function of reduced parallax. This leaves variability as the biggest problem to be dealt with and use of input areas at least 20 mm across in x and y should contain the vast majority of intended inputs for the range of target sizes most likely to be used.

References

Beringer, D. B. and Bowman, M. J. (1989). Operator biases using high-resolution touch input devices. Proceedings of the 33rd Annual Meeting of the Human Factors Society, 320-322.

Beringer, D. B. and Peterson, J. G. (1985). Underlying behavioral parameters of the operation of touch-input devices: Biases, models, and feedback. Human Factors, 27(4), 445-458.

Beringer, D. B. and Lee, K. M. (1988). Parameters influencing the airborne application of touch devices for data entry and retrieval - Phase 1: Technological, positional and chronological effects. New Mexico State U., Human Factors Research Laboratory, Technical Report HFRL-88-2 / MDHC-88-1.

Beringer, D. B. (1989). Touch-panel sampling strategies and keypad performance comparisons. Proceedings of the 33rd Annual Human Factors Society Meeting, 71-75.

Gaertner, K. P. and Holzhausen, K. P. (1980). Controlling air traffic with a touch sensitive screen. Applied Ergonomics, 11, 17-22.

Hall, A. D., Cunningham, J. B., Roache, R. P., and Cox, J.W. (1988). Factors affecting performance using touch-entry systems: tactual recognition fields and system accuracy. Journal of Applied Psychology, 73(4), 711-720.

Karat, J., McDonald, J. E. and Anderson, M. (1986). A comparison of menu selection techniques: Touch panel, mouse and keyboard. International Journal of Man-Machine Studies, 25, 73-88.

Schulze, L. J. H. and Snyder, H. L. (1983, October). A comparative evaluation of five touch entry devices (Tech. Report No. HFL-83-6). Blacksburg, VA: Virginia Polytechnic Institute and State University, Dept. of Industrial Engineering and Operations Research.

Whitfield, D., Ball, R. G. and Bird, J. M. (1983). Some comparisons of on-display and off-display touch input devices for interaction with computer generated displays. Ergonomics, 26(11), 1033-1053.

Navigating Channels Using Parallax Range Lights

Kevin Laxar, PhD
Naval Submarine Medical Research Laboratory
Marc B. Mandler, PhD
U.S. Coast Guard Research and Development Center

Abstract

To determine the navigation performance of observers using current parallax (two-station) range lights, we measured observers' ability to detect deviation from the channel centerline and motion across the channel. We found that the two-point fixed range light configuration affords a reasonably high degree of sensitivity in determining lateral position and motion. With two-point flashing range lights, sensitivity is slightly decreased and errors in judging direction of motion are significantly higher than with the fixed lights. Training or experience was shown to improve navigation performance, and judgments of motion toward the range axis were significantly more sensitive than judgments away.

At the present time, the U.S. Coast Guard employs a visual method for indicating to a vessel's operator the correct path or "range" to follow while proceeding along certain navigation channels, such as approaches to harbors and within rivers. This is the parallax range indication, which for nighttime use consists of a pair of lights positioned on the range axis with the farther light slightly higher than the nearer one (Figure 1). Vertical alignment of the lights indicates that the vessel is positioned on the range's longitudinal centerline, or range axis, and any deviation from this course is readily apparent. As a baseline for comparing performance of other proposed types of range displays, experiments were conducted to measure observers' ability to judge their motion toward or away from the range axis (dynamic simulations), and whether they were on or off the range axis (static simulations).

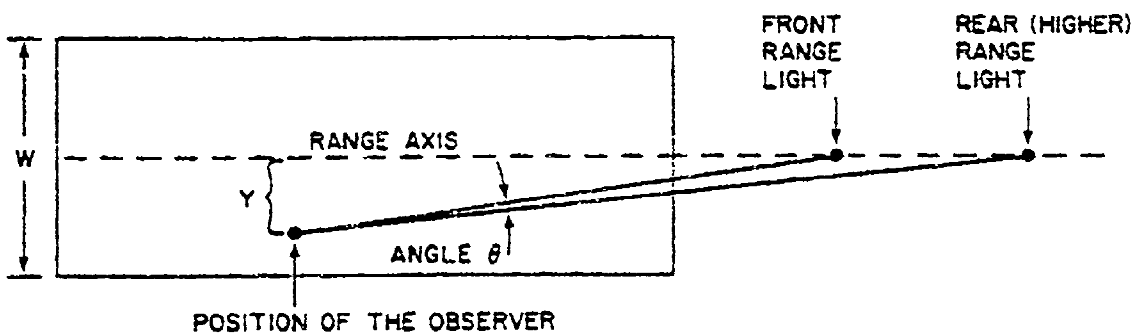


Figure 1. Top view of a parallax range. W: channel width; Y: distance of observer from range axis; θ: horizontal angle between the lights.

Method

Observers

Thirteen observers, 11 men and two women, ages 23 to 59 years, participated. All had 20/25 or better visual acuity, with correction if required, and varied in their experience as psychophysical observers. Four of these observers also participated in the static simulations.

Apparatus

The range configurations were simulated on a Ramtek 9400 high resolution color display system driven by a DEC VAX minicomputer. Observers responded by means of an auxiliary key pad.

Displays

Two types of parallax range indicator lights were simulated:

- Two-point fixed, consisting of two lights which are always on and which are vertically aligned when viewed from the center of the channel. This was simulated as a pair of lights 0.6 arc min in diameter, separated by 4.0 arc min when aligned. When viewed from off center, the two lights are not vertically aligned and the misalignment increases with increasing distance from the center of the channel.

- Two-point flashing, similar to the above except that the two lights are flashed. The upper light exhibited an Equal Interval 6.0 flash characteristic (3.0 sec on and 3.0 sec off), while the lower light showed a Quick Flash characteristic of 0.3 sec on and 0.7 sec off.

Procedure

Observers were dark adapted for 5 min, seated in a dark room 6 meters from the display monitor. The monitor screen subtended visual angles of 2.4° high x 3.3° wide and was uniformly illuminated to 0.003 cd/m^2, equivalent to the night sky with a partial moon. The stimuli were centered on the screen, white in color, and had a luminance of 100 cd/m^2.

Dynamic simulations. For each trial, a pair of range lights was displayed in a configuration corresponding to a view from some distance off the range axis. After a variable foreperiod, the bottom light began to move slowly to the right or left, simulating a vessel's motion across the channel. As soon as the observer could correctly judge the right-left direction of motion, he/she pressed a button corresponding to that direction. When the correct button was pressed the distance traversed by the lower light was recorded by the computer. Trials were separated by a two-sec interval, and errors were recorded and rerun later in the session.

Eleven starting positions were chosen, up to 6.2 arc min right and left of center. The lower light moved at 9.3 arc sec/sec, which was imperceptibly slow so that judgments were based on the position of the lights at some time after the start of the motion. For typical channel configurations, this corresponded to a speed of 2.6 to 11.5 knots across the channel. Performance was measured in a single experimental session, which consisted of one trial at each starting position presented in random order in both left and right directions of motion, repeated over three blocks. The session thus comprised 66 trials, and lasted about 50 min. Observers were given 42 practice trials prior to data collection.

Static simulations. The two lights were presented in one of nine configurations with the lower light up to 37.1 arc sec (0.62 arc min) right or left of the range axis. They were presented in random order once every 4 sec for 0.2 sec. The observer was required to press one of two buttons on the keypad corresponding to the left or right relative position of the lower light. Each position was presented randomly 30 times in two 270-trial sessions which lasted 18 min each, and the computer recorded each response.

Results

Dynamic Simulations

Figure 2 shows the average thresholds for detecting a deviation both left and right of start position for both the fixed and flashing range displays. Threshold is the average deviation from start position required by the observers to correctly judge the direction of motion for that range. The mean threshold for the flashing display is 0.12 arc min higher than for the fixed, but this difference was not significant.

A repeated measures analysis of variance (ANOVA) was computed on the thresholds of each display type for the following factors: 2 Directions of Motion (to the right or left) x 11 Start Positions x 13 Subjects. The direction of motion effect was not significant, but start position had a significant effect on thresholds for both range displays (Fixed: F (10,120) = 4.43, $p < .001$; Flashing: F (10,120) = 3.45, $p < .001$). Thresholds are smallest for start positions at or near the range axis (start position of 0.0) and increase as the start position distance increases left or right from center. This means that observers can easily determine whether they are moving toward or away from the range axis when near the axis, but they require a greater change in lateral position to correctly judge their direction of motion when off the range axis.

The interaction of direction of motion with start position defines the direction of relative motion (DRM) effect, toward or away from the range axis. This interaction was significant for both types of range displays (Fixed:

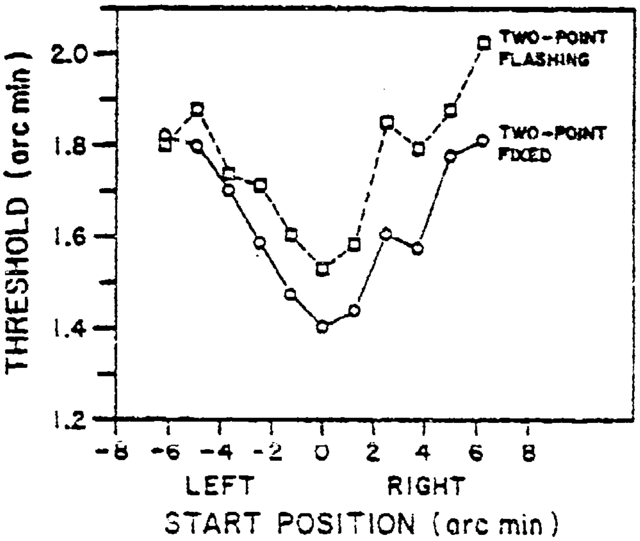


Figure 2. Motion thresholds for two parallax range light displays.

$F\ (10,120) = 7.71,\ p < .001$; Flashing: $F\ (10,120) = 7.36,\ p < .001$), indicating that thresholds for judging motion toward the range axis are different from thresholds for motion away from the range axis. Figure 3 shows these results for each range display. Observers were better at judging changes when the direction of relative motion was toward the range axis than when it was away.

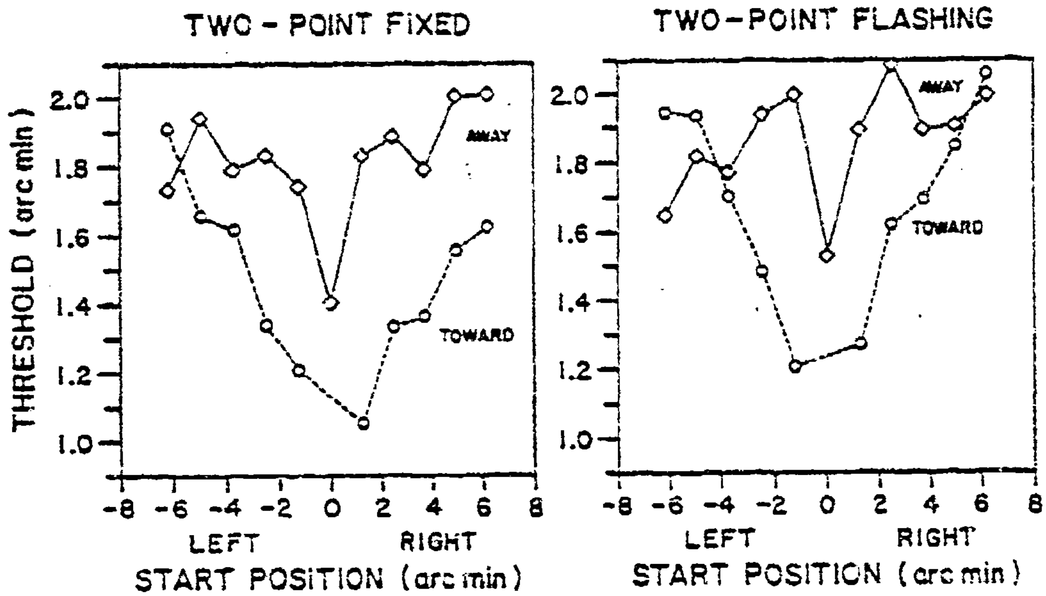


Figure 3. Thresholds for relative motion (DRM) toward and away from the range axis for the two-point fixed and flashing range displays.

Errors--that is, when the observer responded with the wrong direction of motion--were analyzed in a corresponding manner to that for motion thresholds. For the fixed display, mean error percentage was 11.1%, and for the flashing display 17.5%. A Newman-Keuls test showed these were significantly different ($p < .05$). The ANOVA showed that the effect of start position was also significant. Best performance was near the on-axis position, and errors increased with distance off axis.

Static Simulations

Data from the four observers were combined and probit analyses were conducted on the 2160 trials from the two-point range display. With chance performance represented by the 50% probability level and certainty represented by 100%, a probability level of 85% correct was chosen for the practical purposes of this study. Observers could judge when they were off the range axis by 18.4 arc sec (0.31 arc min). Additional practice and a less conservative criterion probability level would likely have made the performance of these observers approach the 5 to 10 arc sec acuity found by Westheimer and McKee (1977) in similar experiments.

Discussion

To relate the measured deviation thresholds to accuracy of navigation it is necessary to convert the angular measures to distances, which depend on the length and width of the range and the placement of the two range lights. The

Commandant, U.S. Coast Guard (1980) has specified optimal limits for such range configurations, so that if a channel width of 152 m (500 ft) is assumed, for example, we can see the navigation accuracy afforded by the range displays.

When on the range axis, the motion threshold for the fixed range lights is 11.9 m at the end of the channel nearest the lights. At the far end of the channel, the accuracy drops to 35.6 m before the navigator can accurately perceive lateral direction of motion. Out near the edge of the channel, performance drops off somewhat, as shown in Figure 2. The motion threshold is 15.3 m at the near end and 45.7 m at the far end of the channel. This is opposite of what is desirable in a range display, which should afford more sensitivity as the vessel approaches the edge of the channel. For the flashing range display, accuracy is somewhat poorer than with the fixed lights. For channels wider than 152 m, the accuracy is proportionally less.

The lateral position threshold determined in the static experiment appears to be substantially more accurate than the motion thresholds. The equivalent accuracy for perceiving which side of the range axis the navigator is on is 2.6 m at the near end of a 152 m wide channel, and 7.9 m at the far end. This simply means that it is much easier for the navigator to tell if he is on or slightly off the range axis than to discern whether he is moving slowly toward or away from the axis. Methodological differences, however, make this task quite different from the dynamic task, so that the difference in accuracy may be overstated.

Four of the 13 observers in the dynamic experiments had extensive experience in making fine perceptual judgments. To determine if such experience had any effect on motion thresholds, we compared their performance with that of the entire group. The experienced observers had thresholds averaging 0.5 arc min more sensitive than the entire group, corresponding to 4.3 meters better accuracy at the near end of a 152 m wide channel, and 12.7 m at the far end.

These results describe the sensitivity afforded by present range light configurations and will serve as a baseline to allow comparison of proposed single-station range lights for evaluating their adequacy as navigation aids. Studies of three single-station range lights proposed by the U.S. Coast Guard are in progress. These findings will be reported in subsequent presentations (Mandler, Laxar, and Luria, 1990) and NSMRL reports

References

Commandant, U.S. Coast Guard. (1980). Range design (COMMANDANT INSTRUCTION M16500.4). Washington, DC: Author.

Mandler, M. B., Laxar, K., & Luria, S. (1990, June). Evaluation of range sensitivity: Comparisons of single-station and parallax ranges. Paper to be presented at the 1990 Conference of the International Association of Lighthouse Authorities, Veldhoven, Netherlands.

Westheimer, G., & McKee, S. P. (1977). Spatial configurations for visual hyperacuity. Vision Research, 17, 941-947.

Prediction of Vocal Performance
Using a Complex Computer Game

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

A complex computer game was adapted to be played using either manual or vocal input. Twelve male and twelve female subjects played ten games, alternating between these two control modalities. Initial vocal performance, gender, and familiarity with computer games were the strongest predictors of subsequent vocal performance. Initial vocal performance was a much better predictor of subsequent vocal performance than was initial manual performance. Familiarity and gender were both stronger individual predictors of performance than either initial vocal or manual performance.

Many careers attempt to select "appropriate" personnel. Adequate selection procedures do not yet exist in a number of these areas; one important example involves air traffic control. Research is being conducted to uncover the best predictors of job behaviors in both the cognitive and non-cognitive domains for this complex task (University of Oklahoma et al., 1988). Presently, only paper and pencil tests are used to screen applicants, even though air traffic controllers' tasks are predominantly vocal (University of Oklahoma et al., 1988). This paper addresses the viability of developing a voice control selection task.

The speech output channel is only one response option for manipulating complex systems, but performance in this modality determines job-effectiveness of air traffic controllers. In other multiple-task environments speech responses can also affect overall performance (Vidulich, 1988). Limited research has been conducted on the differences between vocal and manual control of a task or on factors that predict vocal performance. Wickens' (1984) multiple-resource model separates response modes into vocal and manual categories, suggesting the two rely on different resources, thus having different characteristics. Also according to this, tasks will be best performed when the stimulus, central processing, and response are all compatible (Wickens et al., 1983). This model suggests that spatial tasks controlled by vocal commands will be performed somewhat less efficiently than those controlled manually.

Air traffic controllers use vocal commands to change aircraft flight paths. The job requires this, in spite of the multiple-resource model's predictions of decrease4d efficiency. Paper and pencil tests are still the only tasks used to select individuals for entry into the air traffic control field. The people who perform well on these selection tasks may not necessarily those who would perform best as air traffic controllers. Including a vocal command task as part of a selection battery should enhance the validity of selection.

Although there is little direct empirical evidence, Wickens' (1984) multiple-resource model suggests that performance on a vocal control task rather than a manual one should be a better predictor of subsequent vocal control. Experience and motivation also affect performance. One study on motivation (Dweck, 1986) found that high achieving boys form learning goals while high achieving girls form performance goals. The type of goal can affect performance on difficult tasks. People with learning goals consider challenges an opportunity to learn, while those with performance goals consider them an evaluation with the strong possibility of failure. Subjects' familiarity with the task also can affect the amount of challenge; lack of experience can cause a task to appear more complex than it actually is, thus discouraging those with performance goals. For this reason less experience with computers may lead to lower scores, especially in females. Selection of a computer game for the experiment stems from research that indicates computer games increase motivation and provide accurate, objective, yet unobtrusive data (Gopher, 1988, Porter, 1986).

This experiment was designed to examine differences between vocal and manual task performance and to identify those variables that best predict vocal performance. Vocal control should bias performance toward the more complex task and manual control should increase efficiency on the simple, uncertain task, according to Wickens' (1984) model. Prediction of subsequent vocal performance should be more accurate using previous vocal control performance rather than manual control performance. General experience with computer games should also be a strong predictor due to subjects' development of general metacognitive schemata. Different goals of male and female subjects could make gender another important performance predictor (Dweck, 1986).

Method

Twenty-four cadets (twelve male and twelve female) from the US Air Force Academy volunteered for the experiment. Their ages ranged from 18 to 21. None of the subjects had previous experience with the computer game used. Subjects played the Whale Game using both manual (keyboard) and vocal control. The Covox voice recognition system was used for voice control.

The game required the subject to maneuver a whale around the screen to accomplish two different subtasks. One was to "eat" a mass of plankton that moved across the screen in a random pattern. The second subtask was to cause kayaks that came onto the screen from the boundary and pursued the whale, to crash into stationary icebergs. Both tasks were worth an equal number of points. Points were deducted each time a kayak hit the whale.

During manual control trials subjects used four keys, the Q, E, 2, and S to change the whale's direction. For voice trials, subjects used the spoken letters U, D, L, and R indicating up, down, left, and right. The manual version of the game was modified to approximate the characteristics of the vocal version. The game's pace was adjusted to make manual and vocal games the same length (characters moved about every 1.5 seconds). Pilot testing indicated a voice recognition rate of about 90%, thus a random 10% error rate was incorporated into the manual version. The plankton and kayaks behaved the same in both versions.

Testing took place individually during a one hour and fifteen minute block. Subjects were read a set of instructions and then watched a demonstration of the manual version. Each subject then encoded their pronunciation of the letters U, D, L, and R; all were trained to a minimum recognition criteria before proceeding. Each subject then played ten three-minute games, alternating between voice and manual control.

Results

The data allowed comparisons between voice and manual control versions for both subtasks. Manual performance on the plankton task was significantly superior to voice performance ($t(23)=4.67$, $p<.05$). The kayak scores on the two control versions did not differ significantly.

Regression analyses indicated the two versions had similar task structures (Figure 1). The relationships between task components did not differ significantly for manual and vocal games.

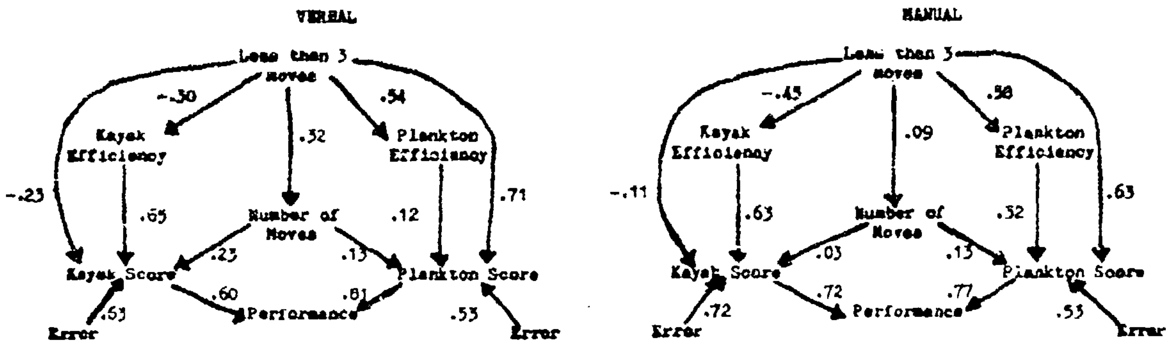


Figure 1: Causal Relationships Between Performance Factors (Numbers are correlation coefficients)

Voice performance was divided into initial (VPERF1) and subsequent (VPERF2) performance. Several measures of individual differences were regressed against this criteria to identify which were most predictive. The strongest predictors were gender (GEND), familiarity with computer games (FAM), and initial vocal performance; together, these three independent variables explained 50% of the variance in subsequent voice performance.

$$VPERF2 = -.34(GEND) + .29(FAM) + .26(VPERF1)$$

$$R^2 = .500 \qquad F = 6.68 \qquad p < .005$$

In contrast, when initial manual performance replaced initial verbal performance in the equation above, its beta weight was only .14 and the overall portion of the variance explained fell by over 4% ($R^2 = .459$). No significant interactions were found between predictor variables. This suggests that these three main effects were additive.

Discussion

The Whale Game combined two types of subtasks: the plankton task was simple but uncertain and the kayak task was complex but certain (Porter, 1986). Equal priority instructions forced subjects to perform both tasks. Higher plankton scores on the manual version might reflect subjects adjusting their strategies to take advantage of greater task-control compatability. The more uncertain task, eating the plankton, was performed better with manual control than with vocal. It may be easier to respond to uncertain events in the manual modality, and since subjects were able to concentrate on either the plankton or the kayaks this difference was reflected by increased plankton scores.

Regression analyses of the two control versions indicate similar task structures. These structures are the implicit strategies revealed by repeated actions across games and subjects (Porter, 1986). The correlations between implicit strategies and overall performance were not significantly different for vocal and manual control. Similar actions, in either the voice or keyboard game, were associated with higher scores, suggesting that the underlying structure of the two versions was not affected by control modality.

The three primary predictors of voice performance were gender, familiarity with computer games, and initial vocal performance. Initial vocal performance was expected to be a strong predictor, due to practice effects and drawing on the same resources. In the Whale Game subjects were biased toward performance goals, by watching a demonstration game and then recording their scores after each trial. Inducing performance (rather than learning) goals may have decreased scores for females and explained the observed effect of gender. Familiarity with computer games was a separate but strong predictor of overall performance. The additive effect of gender and familiarity predicted 44.5% of the variance of criterion scores. General experience with computer games was more strongly predictive of vocal performance than even previous vocal performance. Perhaps the general metacognitive strategies developed through varied experience are even more valuable than initial specific practice wtith the system.

Initial vocal performance was a stronger predictor of vocal performance than was initial manual performance, as was expected based on Wickens' (1984) multiple-resource model. A task relying on different resources than the criteria (i.e. the manual control task) contributed almost nothing to the prediction of performance after gender and familiarity were considered.

Improved selection could be accomplished by including vocal as well as manual control tasks. Spatial vocal tests may be more reflective of general processing resources which, in turn, are more predictive of general capabilities. Incorportating such tasks into the selection battery would increase the probability of selectiong the best personnel. This could lead to reduced training costs, time, dropout rates, and improve subsequent performance. These benefits apply not only to air traffic controllers, but also to pilots and others who must perform some vocal control tasks.

This study has taken an initial step toward discovering more accurate predictors of performance on vocal tasks. It indicates there are performance but not necessarily structural differences between manual and vocal control versions of the same task. The hypothesis that vocal performance would be a better predictor of a vocal task than manual performance was supported. Analysis also suggests metacognitive strategies and general experience can be more important than initial performance when predicting computer game performance.

References

Dweck, C. (1986). Motivational processes affecting learning. American Psychologist, 41, 1040-1048.

Gopher, D., Weil, M., Bareket, T., & Caspi, S. (1988). Fidelity of task structure as a guiding principle in the development of skill trainers based upon complex computer games. Proceedings 32nd Annual Meeting of Human Factors Society. Anaheim, CA.

Porter, David B. (1986). Differential effects of memory load on the performance of substantively different aspects of a computer game. Proceedings of Psychology in the Department of Defense Tenth Symposium (Tech. Rep. No. 86-1) 334-338.

University of Oklahoma, HumRRO International, & Northern NEF (1988). Development of a New Air Traffic Controller Screening Program (Contract No. DTFA-02-85-D-85047). Oklahoma City: University of Oklahoma.

Vidulich, M. A. (1989). Speech responses and dual-task performance: Better time-sharing or asymetric transfer? Human Factors, 30, 517-529.

Wickens, C. D. (1984). Engineering Psychology and Human Performance. Columbus: Charles E. Merrill Publishing.

Wickens, C. D., Sandry, D. L., & Vidulich, M. A. (1983). Compatibility and resource competition between modalities of input, central processing, and output. Human Factors, 25, 227-248.

Authors Notes

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

The authors gratefully acknowledge the support of the Frank J. Seiler Research Laboratory, US Air Force Academy.

Cost-Effectiveness of Home Study using Asynchronous Computer Conferencing for Reserve Component Training[1,2]

Ruth H. Phelps, Ph.D.
Major Robert L. Ashworth, Jr.
U.S. Army Research Institute for the
Behavioral and Social Sciences
Heidi A. Hahn, Ph.D.
Idaho National Engineering Laboratory

Abstract

The resident U.S. Army Engineer Officer Advance Course was converted for home study via asynchronous computer conferencing (ACC). Students and instuctors communicated with each other using computers at home, thus creating an "electronic classroom". Test scores, completion rates, student perceptions and costs were compared to resident training. Results showed that: ACC performance is equal to resident and costs are less than resident.

Geographical dispersion, limited training time and civilian job and family demands make travel to resident schools for training and education difficult for the Reserve Component (RC). Not only is it a hardship for soldiers to leave jobs and family, but their units are unable to conduct collective training when soldiers are absent. In addition, training soldiers at resident schools has become so costly that HQ TRADOC has proposed a 50% reduction in the number of soldiers traveling to resident training by 2007 (TRADOC PAM 350-4).

The purpose of this paper is to summarize an investigation of an alternative means for meeting the educational requirements of the RC. The goals are to (1) develop and test a new training option, using asynchronous computer conferencing (ACC), that

---

[1] These data are summarized from Hahn, H., Ashworth, R., Wells, R., Daveline, K., (in preparation). Asynchronous Computer Conferencing for Remote Delivery of Reserve Component Training (Research Report). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

[2] This paper is not to be construed as an official Department of the Army document in its present form.

would not require soldiers to leave their homes and units and yet maintain the quality of training typically found at the branch school; (2) determine the cost-effectiveness of developing and operating the ACC alternative.

Asynchronous computer conferencing is a means for communicating from different locations at different times (i.e., asynchronously) using a computer network. For training purposes, an "electronic classroom" is established by connecting all students with each other and the instructional staff. A student or instructor can participate in the classroom from any location using existing telephone lines and a computer equipped with a modem. Students can work together in groups, ask questions of the instructors, tutor their classmates or share their thoughts and experiences. Instructors can direct individual study, conduct small group instruction, answer questions, give remedial instruction and provide exam feedback to the students.

Method

Participants

Fourteen RC officers (13 males; 1 female) took Phase III of the Engineer Officer Advanced Course (EOAC) by ACC homestudy. For comparison purposes, performance data were collected from RC students taking the same course in residence at the U.S. Army Engineer School from October, 1986 to June, 1989.

The instructional staff consisted of a civilian full-time course manager/administrator responsible for the overall operation of the course and three part-time instructors. The part-time instructor responsibilities included directing group discussions, remedial instruction and/or monitoring student progress.

Course Description

Course materials consisted of Module 6 of the EOAC (66 program hours of instruction). Media used included paper based readings and problems, computer-aided instruction, video tapes and computer conferencing discussion. Topics covered were Army doctrine (e.g., rear operations), technical engineering (e.g., bridging, flexible pavements), leadership and presentation skills. The program of instruction was identical for the ACC and resident classes.

Equipment, Procedure and Data Analysis

Each student was provided with an IBM XT computer with 20 megabyte hard disk, color monitor and printer. Software and courseware loaded on each computer consisted of: (1) a specially developed course management system and communications package; (2) computed-assisted instruction and tests; (3) word processing package; (4) spreadsheet.

Communication software for asynchronous computer conferencing was provided through U.S. Army Forum, Office of the Director of the Army Staff. The host computer was located at Wayne State University and used the CONFER II conferencing software system.

The course was conducted from September, 1988 to April, 1989. Students were mailed all their computer equipment with written assembly and operation instructions and course materials. In addition they were provided with a toll free "hot line" telephone number for resolving hardware/software problems. The first lessons to be completed were self-conducted and designed to familiarize the student with the operation of the computer and software. Scores for computer training were not included in overall course grades.

Part-time instructional staff were provided the same equipment and software as the students. In addition they were given a 40 hour training course on operating the hardware/software, instructional responsibilities and teaching/motivational techniques. Instructional staff and researchers met together to conduct this training using a combination of lecture and hands-on practice with the computer.

There were four types of data collected: (1) test, practical exercise and homework scores; (2) pre- and post course student perceptions of their amount of knowledge on the course topics; (3) course completion; (4) cost of converting and executing the course. Comparisons of the resident to the ACC course were made using multivariate analysis variance procedures for a two-group design.

Results

As shown in the top of Table 1, there was no reliable difference between the test scores of students in residence versus ACC. A comparison of the students' self ratings of their level of knowledge before and after the course, showed that the ACC group had significantly greater gains in their perceived amount of learning, as shown in the bottom of Table 1. Completion data showed that 95% of resident students completed the course compared to 64% of the ACC students.

Table 1

Student Scores and Ratings

Scores	ACC	Resident	Significance
Tests	92.0%	86.4%	NS
Homework	88.8%	92.0%	NS
Practical Exercise	90.4%	89.9%	NS
Perceived Amount Learned (% Post-Pre)	33%	12%	$p<.05$

Cost data were computed separately for (1) converting an existing course for delivery by ACC and (2) executing each iteration of the course. If the conversion were done by within-government staff, then the cost would be approximately $296,100. If it were done under contract, then the cost is estimated at $516,200. Start-up costs of equipment purchase and instructor training were estimated to be $73,100 for within-government and $96,000 for contractor. Costs that will recur with each iteration were estimated at $234,400 for within-government and $420,900 for contractor.

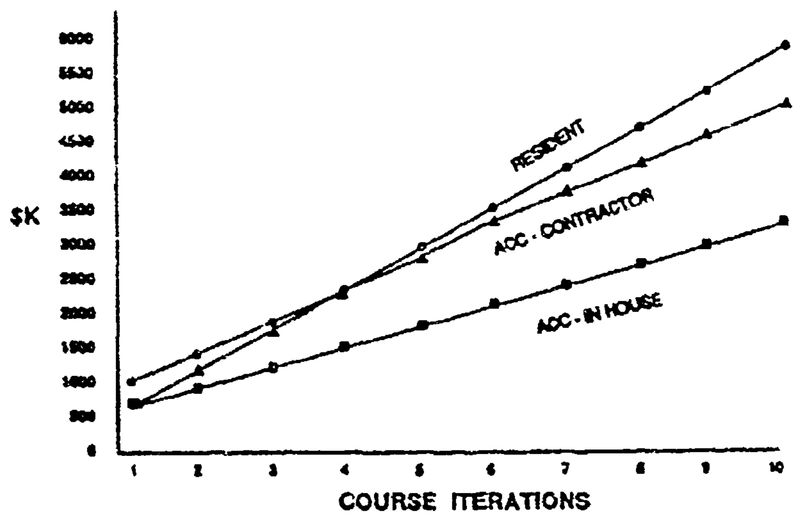


Figure 1. Relative costs of EOAC alternatives over 10 course iterations.

Figure 1 shows the total course conversion, start-up plus the recurring costs over 10 course iterations. Initially resident and ACC (within government) are similar with ACC (contractor) costs being nearly twice as much. However, when the costs of conversion and execution are amortized, ACC (contractor) becomes less costly than resident training after four course iterations. After five iterations ACC (within government) would save 47% and ACC (contractor) would save 6%.

Cost-effectiveness ratios were computed by combining the cost and completion rate data. The ratio was greatest for ACC using government staff (.64), second for resident training (.41), and lowest for ACC using contractor staff (.36).

Discussion

It has been shown in this report that there is a cost-effective alternative to sending RC soldiers to branch schools for resident training. Training by ACC can be conducted just as effectively and for less money. Thus, this technology appears to meet the need of the RC to complete educational requirements from the home or homestation, without long absences from the unit. The "electronic classroom" could be conducted remotely from existing educational institutions such as the branch school and/or the U.S. Army Reserve Forces School in order to maintain standardized instruction.

Additional research is needed, however, to improve the completion rate for ACC home study. Reasons for dropping out of the experimental course were related to limited time due to competing activities such as civilian jobs and family. A means of predicting which soldiers are likely to succeed or drop out of home study will assist Army trainers in both selecting students and providing assistance for those at high risk.

References

U.S. Army Training and Doctrine Command. (1989). Army Training 2007. (TRADOC Pamphlet 350-4). Ft. Monroe, VA: Author.

College Characteristics and Grade Point Average

Jacobina Skinner
Stephen D. Armstrong
Air Force Human Resources Laboratory

Abstract

Past research has demonstrated a relationship between undergraduate grade point average (GPA) and training and job performance. Further evidence suggests that the undergraduate institution attended may play an important role in influencing that relationship. This paper presents a methodology for establishing empirically the effect of undergraduate institution on the predictive validity of GPA. Procedures are also proposed for exploring the responsible characteristics of undergraduate institutions if college effects are detected. Implications for personnel selection systems of differential GPA validity for colleges are discussed.

Throughout the military and private sector, undergraduate grade point average (GPA) plays an important role in job selection decisions. This measure of academic achievement and demonstrated ability is widely held to predict employee performance. Recent literature reviews show significant but modest relationships between GPA and employee performance both in training and on the job (e.g., Dye, Reck, & Harris, 1989).

An issue raised by the use of GPA as a personnel selection factor concerns possible inequivalencies in the grade scale across colleges. The implication for employers is that expected performance would vary among job applicants who have the same GPA, but graduate from different colleges. Research on this issue is sparse but two studies suggest that a school factor may moderate the GPA-performance relationship. Dye et al. (1989) have found correlations for graduates of the same college to be higher on average than those for graduates of different colleges. Further evidence that college characteristics may influence the predictability of GPA has been reported for Air Force officers commissioned from the Reserve Officer Training Corps (ROTC) program (Barrett & Armstrong, 1989). Performance prediction was improved by considering a quality measure for the officers' college in addition to their GPA.

The current study extends the investigation of the college and GPA issue in the Air Force to a second officer commissioning source: the Officer Training School (OTS) at Lackland AFB, TX. The design and preliminary findings of a two-phase study of the relationship between GPAs awarded to cadets graduating from different colleges and their subsequent performance in OTS are described in this paper. In the analytic phase, the initial focus is on the validity of GPA as a cadet selector. In addition to simple GPA effects, a methodology for examining the joint effect of GPA and college is described. If differential validity for colleges is observed, follow-on analyses have been designed to explore the responsible variables. In the explanatory phase, a primary interest will be to determine the relative contribution of student enrollment standards versus post-admission experiences to the college effect. Procedures are described for determining whether the more important explanatory variables reflect the talent of entering students as opposed to the quality of education offered by the college.

Analytic Phase: GPA and College Relationships with Cadet Performance

OTS cadets are selected using a whole-person approach by boards composed of field grade officers. Previous research has demonstrated that GPA is one of the factors related

($r = .26$, $p < .01$) to the selection board's decision to admit or reject a candidate (Cowan, Barrett, & Wegner, 1989). Other significant factors included the Air Force Officer Qualifying Test (AFOQT), degree type, and prior military service record.

One goal of the analytic phase was to evaluate the validity of the GPA selector by obtaining an accurate estimate of the population correlation with OTS performance. Because sample size affects the reliability of the estimate, procedures were developed to maximize available cases. Further, to account for sample homogeneity produced by prescreening cadets on GPA and AFOQT, restriction in range corrections were applied to the correlation coefficients.

A separate series of analyses was designed to address the comparability of the GPA-performance relationship for different colleges. Graduates from colleges nation-wide are considered by OTS boards, but from many of the colleges there are only a few applicants. To ensure that within-college sample sizes were large enough to detect effects reliably, it was necessary to restrict the total sample to selected colleges.

Method

Procedure

Data were obtained from archival personnel files maintained on Air Force officers. A total sample of 11,619 cadets who entered OTS between 1982 and 1988 was identified. Source data for the primary predictor variables were the cadets' 4-year undergraduate GPA anchored on a 4.0 scale and the college which conferred their baccalaureate degree. Measures of cadet performance were obtained from various phases of the 12-week OTS program. Information on reason for terminating training was used to generate a Pass/Fail dichotomy reflecting final training outcome for the total sample. Eight additional measures of performance were available for graduates ($N = 9{,}858$). Final Course Grade was an overall rating of academic success in the training course obtained by averaging scores on five Consolidated Written Tests (CWT 1 - 5). The CWT measures were percent correct scores achieved on exams covering four basic areas: defense studies, professional military knowledge, communication skills, and leadership and management principles. The final two measures were ratings by training instructors of the cadets' overall accomplishments. The Officer Training Effectiveness Report ratings were reported on a 5-point scale (Outstanding = 4, Excellent = 3, Satisfactory = 2, Marginal = 1, Unsatisfactory = 0) at the 6^{th} and 11^{th} week of training. Each OTER considered performance areas including leadership, adaptability to military life, professional qualities, communication, and judgment.

Analysis

Simple GPA Effect. Summary statistics were obtained to describe performance on the GPA predictor and training criteria. Simple (bivariate) correlations to describe GPA-performance relationships were also computed. To account for restriction in range on the AFOQT, a multivariate correction was applied using test results for an unrestricted sample of 3,000 examinees (Skinner & Ree, 1987). For GPA, the unrestricted group consisted of OTS applicants ($N = 32{,}648$) from 1981 to 1988.

Joint GPA and College Effect. To examine the joint effect of undergraduate institution and GPA on cadet performance, a reduced sample was identified. Subjects were restricted to graduates of colleges which had conferred degrees upon 20 or more cadets. A case count of 20 per college was judged to be the minimum necessary to detect college effects reliably. A total of 5,150 cadets from 102 colleges met the requirement.

Multiple regression analyses are currently being conducted to test a priori hypotheses, following the generalized linear model approach described by Ward and Jennings (1979). The starting model is designed to insure a relatively complete specification of potential joint effects due to GPA and college. College identity is measured using binary coding to form categorical membership predictors for the 102 institutions. In addition, the starting model contains first-, second-, and third-order polynomial terms for GPA and their interactions with the college predictors. Nonlinear GPA predictors are included to account for a possible relationship with expected performance that is ogival in shape.

Reduced models have been designed to detect other potential relationships for GPA and colleges with cadet performance that are less complex than the one hypothesized by the starting model. Possible outcomes are an interaction between GPA and college, but one that has a simpler functional form, either linear or curvilinear. Alternatively, there may be a joint but noninteracting effect due to GPA and college (with either a linear, quadratic, or cubic form). In these cases, expected performance would differ by college at fixed GPA values, but the difference per unit change in GPA would be constant. The least complex alternate outcome would be an effect due solely to GPA (linear, quadratic, or cubic) or solely to college.

To isolate the "best" model, pairs of models will be compared using the F-test ($p < .01$). The same series of analyses are planned to identify the most appropriate model for each criterion. Raw score weights derived from the least-squares solution will be used to compute predicted performance scores. These results on the magnitude and direction of significant effects will aid in interpreting, for example, whether any differences in expected performance for equivalent GPAs as a function of college are appreciable.

Explanatory Phase: Characteristics Which Account for College Effects

The explanatory phase will be accomplished if results of the analytic phase show that the relationship between GPA and cadet success varies by college. College effects will be judged to be present for a criterion if the corresponding "best" model contains significant information about college membership. The objective is to identify variables underlying the effect attributable to colleges. A primary interest is whether performance variance accounted for by colleges is due principally to admission standards (college selectivity) or to the nature of academic experience (educational quality). The college quality indicator found to enhance the predictability of ROTC cadet performance (Barrett & Armstrong, 1989) incorporated both components in a single measure using policy-specification analysis. The proposed design provides procedures for partitioning the college selectivity and educational quality components in order to determine the unique contribution of each to the college quality factor. Further, provisions are made to identify specific properties of the academic and instructional opportunities offered by colleges that may account for an educational quality component.

Method

Procedure

The unit of analysis will be colleges, as specified in the analytic phase ($N = 102$). Measures of the college selectivity and educational quality components are being obtained from published reports which rate and describe colleges (e.g., American Council on Education, 1987). To measure college selectivity, average scores of the entering freshman class on national standardized aptitude tests (Scholastic Aptitude Test, American College Test) are being recorded. Several potential measures of educational quality have been identified. Ratings of overall academic excellence and academic-athletic balance are available from a survey of college administrators and faculty (Gourman, 1985). Other measures are percentage of applicants accepted, percentage of graduate students, ratio of students to full-time faculty, percentage of full-time faculty with PhDs, number of volumes in library, and dollar value of endowments, grants, gifts, and donations. Data to be used as criteria reflect the unique contribution of the 102 colleges to the prediction of OTS cadet performance. These values are the regression weights (b-weights) for the college membership variables from the "best" model in the analytic phase.

Analysis

Regression analysis will be used to explore the relative contribution of college selectivity and educational quality measures in accounting for the college effect. The b-weights for colleges will be regressed on college selectivity variables only, educational quality variables only, and both. Predictive accuracy will be compared using R^2 and F statistics. Stepwise procedures will also be accomplished to identify the most salient indicators among the available educational quality measures.

Results and Discussion of GPA Validation

Descriptive statistics of the performance criteria in Table 1 show that most cadets passed the training program (86%) and achieved mean scores in the low 90s on the Final Course Grade measure and on the CWTs. On the average graduates received satisfactory ratings on the 6th week OTER and improved their performance to an excellent rating by the 11th week. Mean GPAs for OTS applicants, entrants, and graduates showed a slight increase across groups (2.96 to 3.05), but the standard deviation values did not change (.45 to .44).

As shown in Table 1, the uncorrected correlations between the performance criteria and GPA indicated low to medium-low positive relationships. The lowest correlation was observed for the Pass/Fail dichotomy (r = .01) and the highest correlation for Final Course Grade (r = .31). All correlations were significant at $p < .01$ except Pass/Fail and the 6th week OTER. Correlations corrected for restriction in range on both GPA and AFOQT were not appreciably larger (.03 or less) than the uncorrected correlations.

Results of the preliminary GPA validation are consistent with earlier studies (e.g., Cowan et al., 1989) showing the utility of GPA as a factor in the OTS selection process. For many criteria of cadet performance, the observed relationships with GPA are statistically significant and are judged to be appreciable as well. The highest gains in expected

Table 1. Summary Statistics of OTS Criteria and Correlations with GPA

Criterion[a]	Mean	SD	r	r_c[b]
Pass/Fail	.86	----	.01	.03
Final Course Grade	91.78	3.78	.31**	.33
CWT 1	91.59	5.24	.19**	.21
CWT 2	91.57	5.18	.22**	.25
CWT 3	91.84	5.06	.21**	.24
CWT 4	90.90	5.46	.22**	.24
CWT 5	92.39	4.85	.18**	.20
OTER 6th Week	2.04	.68	.07*	.08
OTER 11th Week	3.78	1.42	.20**	.22

[a]Pass/Fail N = 11,619; other criteria N = 9,858.
[b]r_c = corrected for restriction on GPA and AFOQT.
*$p < .05$
**$p < .01$

performance per one point increase in GPA are for Final Course Grade (about 2.5 points on the grade scale or .75 standard deviation units) and for the 11th week OTER (.7 point on the rating scale or .50 standard deviation units).

The comparable magnitude of the uncorrected and corrected correlation coefficients is an unexpected finding, in view of prior research suggesting that selection board members use GPA information in their evaluations of OTS cadets. Similar corrections of AFOQT validities for the study sample produced marked increases. Apparently, the boards placed greater emphasis on applicant aptitude, and perhaps on degree type (technical or nontechnical), thereby mitigating the anticipated impact of sample curtailment on GPA.

Implications for Personnel Selection

An advantage of the proposed methodological approach is the relative ease of application by other government and private sector agencies. Special data collection efforts, such as convening policy-makers to judge college quality, are not required. Most job applications request the information on college identity needed to accomplish the analytic phase and explanatory data on college characteristics are readily available from published documents. However, a potential limitation is that the accuracy of published college ratings is unknown.

If college effects are detected, personnel selection systems which include GPA should account for the moderating influence of the underlying characteristics of undergraduate institutions. If college selectivity is the most salient characteristic, selection systems which use an aptitude measure in addition to GPA may already be capturing the performance variance. Other systems which rely on GPA exclusively may find college selectivity to be an aptitude surrogate. Either type of system might also be improved by considering indicators of academic and instructional experiences if educational quality is shown to moderate GPA predictability.

References

American Council on Education (1987). American universities and colleges (13th Ed.). New York: Walter de Gruyter.

Barrett, L.E., & Armstrong, S.D. (1989, November). Moderating effects of school characteristics on the predictive validity of college grade point average (GPA). Paper presented at the 31st Annual Conference of the Military Testing Association, San Antonio, TX.

Cowan, D.K., Barrett, L.E., & Wegner, T.G. (1989). Air Force Officer Training School selection system validation (AFHRL-TR-89-65). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Dye, D.A., Reck, M., & Harris, P. (1989, August). Use of college grade point average to predict job performance. Paper presented at the 97th Annual Convention of the American Psychological Association, New Orleans, LA.

Gourman, J. (1985). The Gourman report: A rating of undergraduate programs in American and international universities (5th Ed.). Los Angeles, CA: National Education Standards.

Skinner, J., & Ree, M.J. (1987). Air Force Officer Qualifying Test (AFOQT): Item and factor analysis of Form O (AFHRL-TR-86-68). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Ward, J.H., Jr., & Jennings, E. (1979). Introduction to linear models. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Institutional and Occupational Paths to Retention Among United States Air Force Officers

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

Abstract

This study examined the hypothesis that institutional reasons for joining the Air Force--vice occupational reasons--would predict career retention. Contrary to expectations, survey responses from 9,409 officers demonstrated consistent positive relationships ($p < .05$) between retention and several occupational reasons for joining (particularly job security and benefits). Institutional reasons generally failed to predict retention. Officers with a strong desire to fly or prior flying experience were poor retainers. Policy implications are discussed.

Moskos's institutional/occupational (I/O) thesis of organizational change provides an important framework for understanding many current trends in the armed services. Moskos argues that "the American Military has been moving away from an institutional format to one that increasingly resembles that of an occupation [Mokos & Wood, 1988, p.3]." Institutions are holistic organizations based on service to others and a sense of "calling" or "duty," requiring a near-total commitment to the organization. Occupations, on the other hand, are based on employment practices of supply-and-demand, an exchange of employees' time for money, and a specialist's job orientation.

Moskos and his colleagues offer clear evidence of the gradual transition from an institutionally based to an occupationally based military. Moskos further outlines three consequences of the institutional-to-occupational transition: (a) a tendency for employees to define task boundaries and standards, (b) replacement of intrinsic motivation with extrinsic motivation, and (c) an undermining of military professionalism. The consequences may be severe for the military: Troops given to a cost-benefit analysis of their occupational behaviors may refuse to endure hardship and make the extreme sacrifices sometimes required.

An additional--and generally unexamined--question is the effect of this occupational trend on more routine organizational behaviors: job performance, satisfaction, absenteeism, and retention. The charter of the ATC Officer Selection Study Group gave us a unique opportunity to examine the effects of I/O orientations on retention in the Air Force. The Study Group was commissioned by the ATC senior leadership in October, 1988, and charged with identifying attributes of officer candidates that predict success (with an emphasis on retention) in the Active Duty Air Force. The Group examined a host of personnel system and survey variables to predict long-term retention. Included were eight I/O items that provided a direct test of the effects of officers' initial I/O orientations on their likelihood of retention.

Moskos and Wood imply that institutional commitment should yield higher retention rates than will occupational commitment. They propose that "institutional identification fosters greater organizational commitment and performance than does occupational [pp.4-5]." Watson & Appel (1986) identified organizational commitment as a crucial intervening variable in service members' turnover decisions. It follows that institutionally oriented officers should be our best retainers. However, it may be that officers with a strong occupational orientation, and who find the military a good source of employment (good pay and benefits, security, training, etc.), retain best. Therefore, the best retention may come from either institutionally oriented officers, occupationally oriented officers, or those high on both dimensions.

Method

Subjects

The Study Population File combined elements of existing Air Force personnel files and contained records for 146,565 line officers with a Total Active Federal Military Service Date (TAFMSD) of 1965-84. From this file a survey sample of 15,264 officers was drawn by stratified random sampling. The stratification factors were: Cohort Group (1965-74, 1975-80, and 1981-84), Commissioning Source (USAFA, OTS, and ROTC), and Utilization Area (Pilot, Navigator, Engineer, and Other). A total of 9,409 survey packages (61.6%) were completed and returned in usable form.

Survey Design

The study group constructed the Officer Accession Characteristics Survey to supplement the information in the Study Population File, in order to retroactively measure preselection and precommissioning data as accurately as possible. The survey contained 93 items and covered eight broad areas: family and geographical background, early military exposure, initial entry decision factors, previous employment, high school and college activities (both athletic and nonathletic), key life events, flying interest and experience, and initial military expectations and experiences.

The section on initial entry decision factors included ten items, eight of which tapped the officer's initial I/O orientations. Each respondent was asked to rate (on a five-point Likert-type scale) the importance of each of the following factors in his/her decision to enter the Air Force:

Opportunity to serve my country (Institutional, I)
Opportunity to make the world a better place (I)
Continuing a family tradition of military service (I)
Steady work/job security (Occupational, O)
Good pay (O)
Better job/promotion opportunities than in civilian life (O)
Opportunity for training/education (O)
Good benefits (O)

These items were adapted from those used by Segal and Blair (cited in Moskos & Wood, 1988), rephrased to apply to the officer's initial entry decision. In addition to separate analyses for each item, the three institutional items and five occupational items were averaged to compute an overall institutional measure (Inorient) and an overall occupational measure (Ocorient).

Study Design

The measures of primary interest were the correlations of the separate and combined I/O measures with the criterion of retention. Because of maturation differences, this criterion was defined differently for each cohort group. For the 1965-74 Cohort Group, "success" was defined as retention to 14 years of Total Active Federal Military Service (TAFMS), the point at which the officer attrition curve stabilizes. (Ninety-seven percent of officers who reach the 14-year mark stay to retirement.) Success for the 1975-80 Cohort Group was defined as retention to 8 years (the point at which officers have served their initial commitments, plus at least one year), since these officers have not yet had the opportunity to stay to 14 years. The 1981-84 Cohort Group was deleted from the present analyses but was reserved for future analyses as these officers mature to a minimum criterion of 8 years.

Results and Discussion

Correlations with Criterion

Table 1 presents the correlation coefficients of all eight I/O variables, as well as Inorient and Ocorient, with retention. These results present some surprises. First, while several occupational items (including the composite measure, Ocorient) predict long-term retention, the institutional items generally do not. Only "Join-Country" and "Join-World" show any correlation with retention, and these effects are specific to just two cells of the analysis. The composite measure, Inorient, likewise fails to predict retention for any group other than 1975-80 Pilots. This is surprising in that an orientation of service and military tradition would seem to predict a long-term affiliation with the military service. However, this is not the case.

On the other hand, the effect of occupational orientation is both robust across items and cells and--as our second surprise--positive. The composite measure predicts retention for all nonpilot cells but one. While some items have little predictive power, others ("Join-Security," "Join-Benefits") predict well for most groups. For the nonpilot utilization areas, a strong occupational orientation predicts a propensity to stay in the Air Force.

These findings refute our hypothesis that institutionalists are keepers and occupationalists are leavers. Instead, occupationalists are keepers and the institutionalists show no strong inclination either way. This indicates that nonpilots who enter the Air Force looking for a good job find what they are looking for. Although they do not stay because of institutional reasons such as dedication to duty and love of country, they do stay because the Air Force is a good place of employment.

The consistent positive predictions produced by "Join-Security" and "Join-Benefits" are particularly telling in this regard. Nonpilots (and to a lesser extent pilots) who joined for security and benefits are keepers. This

TABLE 1
Correlations of I/O Variables with Retention*

	1965-74 Cohort Group				1975-80 Cohort Group			
	Eng.	Nav.	Oth.	Pil.	Eng.	Nav.	Oth.	Pil.
INSTITUTIONAL ITEMS:								
Inorient								+.08
"Join-Country"								+.10
"Join-World"								
"Join-Tradition"							-.09	
OCCUPATIONAL ITEMS:								
Ocorient	+.14	+.11	+.16		+.19	+.15		
"Join-Security"	+.25	+.17	+.22	+.11	+.22	+.27		+.06
"Join-Pay"	+.12		+.06					
"Join-Better Than"			+.12					
"Join-Training"	+.16		+.10	-.17	+.14			-.13
"Join-Benefits"		+.12	+.14		+.14	+.19	+.07	

*Only significant ($p < .05$) findings are reported

TABLE 2

Correlations of "Flying Interest" Variables with Retention*

	1965-74 Cohort Group				1975-80 Cohort Group			
	Eng.	Nav.	Oth.	Pil.	Eng.	Nav.	Oth.	Pil.
Join-Training	+.16		+.10	-.17	+.14			-.13
Join-To Fly	-.14		-.15	-.10	-.20	-.17	-.11	-.08
Pilot Hours	-.13						-.11	-.09
Pilot Rating				-.08		-.15		

*Only significant ($p < .05$) findings are reported.

finding implies that the Air Force is an attractive organization in these aspects and is meeting its officers' needs. On the other hand, the low correlations with "Join-Pay" imply pay may be less important than security and benefits as a motivator for long-term retention. The message for Air Force personnel policy seems clear: Initiatives and proposals that threaten security and benefits may hurt retention. Pay issues may be less important.

Contrary to the other utilization areas, pilots showed no correlation between Ocorient and retention for either cohort group (r = .04 and .02 for the respective cohort groups). Tests of significant differences between correlation coefficients indicate that pilots are significantly different from other utilization areas (with the exception of the 1975-80 "Other" officers) in this regard. Thus, occupational factors do not seem influential in either keeping pilots in or driving them out of the Air Force. The one exception to this rule is "Join-Security," which positively predicts pilot retention. Contrary to popular logic, occupationally oriented pilots do not leave the Air Force simply due to the lure of lucrative airline jobs: Ocorient is uncorrelated, rather than negatively correlated, with pilot retention.

This difference between pilots and other officers is explained by the "Join-Training" correlations in Table 1. "Join-Training" is positively correlated for three of six nonpilot cells but negatively correlated for the pilots. Pilots who join the Air Force for training are leavers, while other officers who join for training are (for the most part) keepers.

Table 2 clarifies this trend. Although "Join-Training" is negative for pilots only, "Join-To Fly" and other indicators of flying interest/experience are consistently negative. Officers who join the Air Force to fly or enter with flying experience are leavers--regardless of whether they become pilots. These data point to a curious conclusion: Whereas the Air Force may be a good place to work (as evidenced by the positive correlations with the occupational items), it may not be such a good place to fly.

<u>Correlations Within the I/O Predictor Set</u>

To examine the correlations within the I/O predictor set, we developed four correlation matrices, breaking the survey sample out by the two cohort groups and pilot vs. nonpilot utilization areas. The resultant matrices reveal a consistent and fairly powerful set of positive correlations. Of the 112 pairwise correlation coefficients, 93 are significant ($p < .05$) and positive; only eight are negative (none significant). Correspondingly, the correlations between Inorient and Ocorient are consistently positive:

1965-74 Nonpilots	$r = .23$, $p < .001$
1965-74 Pilots	$r = .23$, $p < .001$
1975-80 Nonpilots	$r = .04$, n.s.
1975-80 Pilots	$r = .12$, $p < .001$

These results clearly refute any notion of a negative relationship or tradeoff between an individual's occupational and institutional orientations. At least for these self-reported reasons for entering the Air Force, there is a positive I/O relationship. These findings match those by McCloy and Clover (1988), which also failed to detect a negative I/O relationship.

These findings warrant replication with different populations, other I/O items, and different behavioral settings. If supported, they have a clear practical application: Far from opposing each other, I and O orientations seem to complement each other. An individual who is serving for (altruistic, duty-oriented) institutional reasons may also serve for (relatively selfish, calculated) occupational reasons. As a consequence, initiatives by senior leadership to curb the rising tide of occupationally oriented "careerism" and replace it with institutional values should be reconsidered.

Implications

These findings have several clear implications, both for the research community and for selection/personnel policy:

1. Occupationally oriented nonpilot candidates will most likely be keepers, particularly those attracted by job security, benefits, and the opportunity for training.
2. Pilot candidates with a strong desire to fly (as evidenced by self-report or prior flying experience) will be poor retainers. Likewise, pilot candidates interested primarily in the Air Force's training opportunities will be poor retainers.
3. Personnel initiatives that threaten benefits and job security will impair retention.
4. Selection and training initiatives that attempt to replace "careerist" orientations with institutional orientations should be reconsidered. Institutional and occupational orientations appear to be complementary, rather than contradictory. Furthermore, an occupational orientation positively predicts higher retention.

References

McCloy, T. M., & Clover, W. H. (1988). Value formation at the Air Force Academy. In C. C. Moskos & F. R. Wood (Eds.), The military: More than just a job? (pp. 129-149). Washington: Pergamon-Brassey.

Moskos, C. C., & Wood, F. R. (1988). Introduction. In C. C. Moskos & F. R. Wood (Eds.), The military: More than just a job? (pp. 3-14). Washington: Pergamon-Brassey.

Watson, T. W., & Appel, V. H. (1986). Implications of the results of recent turnover research for Air Force policy (Report No. AFHRL-TP-86-5). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Author Notes

The Officer Selection Study Group consisted of: Lt Col F. Randall Starbuck, Lt Col Merl A. Morehouse, Lt Col Daniel E. Hancock, Maj William R. Cummings, Maj David K. Roberts, Capt Glenwood M. Barbee, Capt Gary R. Grimes, SSgt Wayne K. Getson, Sgt Phillip M. Harris, and Ms. Ingrid Wright.

Selection for Retention: Report of the ATC Officer Selection Study Group

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

Abstract

This study examined a cohort file of 148,566 officers, with survey responses from 9,409 of these officers, to determine attributes of officer candidates that predict retention in the Active Duty Air Force. Variables that reliably ($p < .05$) predict retention include: source of commission, college cost, military order of merit, reason for joining the Air Force, precommissioning pilot rating, and family background (income, mobility, military background, and socioeconomic status). Regression models succeeded in accounting for 12 to 72 percent of the variance in retention behavior, depending upon the specific group studied.

Careful screening and selection of applicants is a traditional concern of the military services. However, even the most capable candidate may have little value for a retention-oriented force (such as the Air Force) if his/her likelihood of career retention is low. In addition to producing career-oriented warriors, scientists, and technicians, the Air Force frequently becomes a valuable training conduit for civilian industry. This problem has become particularly serious for the pilot force in recent years. It is clearly in the service's interest to access candidates with known retention characteristics--most often those who are likely "keepers."

On the other hand, a priori identification of candidates likely to be keepers is no easy matter. The voluminous turnover literature provides only limited information on preselection predictors. The primary focus of the work on turnover has been postselection predictors (e.g., Gibb, Nontasak, & Dolgin, 1988; Watson & Appel, 1986), although some studies examined preselection predictors for populations very different from officer candidates (Finstuen & Alley, 1983; Ward & Tan, 1985). Furthermore, most studies examined relatively short-term turnover (Cook & Morrison, 1983; Shenk, Watson, & Hazel, 1973) or turnover intent (Dunn & Feller, 1983; Gibb et al., 1988), rather than actual turnover.

The senior leadership of Air Training Command addressed this problem to the Officer Selection Study Group in October, 1988. The Group's charter was to identify attributes of officer candidates that predict success (with an emphasis on retention) in the Active Duty Air Force. We had full access to and cooperation from the Air Force's personnel and scientific communities. Through these resources, we had access to all available Air Force personnel files, as well as the technology to develop new data collection instruments. An important qualification, however, was that our final report was due within seven months, for submission to the Air Force senior leadership. This combination of maximum access to resources, maximum visibility, and minimum turnaround time quickly established the project's direction.

Method

The basic study plan was to examine descriptive statistics and build a multiple linear regression model to predict retention behavior from all available precommissioning attributes of officer candidates. Data sources consisted of existing data files and a survey created specifically for this study. Because the Air Force accesses officers from a variety of backgrounds and employs them in a variety of duty areas, the study was stratified by both commissioning source and utilization category. The study was further stratified by date of entry, since the basic data file covered a wide span of entry dates (1965-88).

Study Population File

The Study Population File combined elements of existing files and contained records for 148,566 officers. These records were extracted from the Air Force Human Resources Laboratory's (AFHRL) Officer Historical Profile, with the following criteria: line officer; commissioned through OTS, ROTC, or USAFA; Total Active Federal Military Service Date (TAFMSD) of 1965-84, and currently on active duty or with a verifiable separation/retirement date. The Officer Historical Profile contained three categories of data available through the personnel system: (a) entry data, such as home of record, source of commission, ethnic group, sex, and marital status; (b) duty history, such as assignment history, promotion record, and Officer Effectiveness Report (OER) data; and (c) separation data, such as separation date and reason.

This extract formed the core of the cohort data base and defined the study's population. The following additional data elements were extracted from supporting files: (a) standardized test scores, including Air Force Officer Qualifying Test (AFOQT), SAT, and ACT scores, as available; (b) USAFA application data, for all Academy graduates; (c) additional personnel system data (elements not available through the Officer Historical Profile), including undergraduate academic institution, AFROTC detachment, and overseas tour data; and (d) college data, to categorize major colleges by cost, selectivity, religious affiliation, size, and athletic conference.

Survey File

Although the supporting files provided a wealth of background data, they were not designed to predict retention. Consequently, the variables from these files were unlikely to account for an appreciable amount of the variance in retention behavior. Therefore, the study group constructed the Officer Accession Characteristics Survey to retroactively measure preselection and precommissioning data as accurately as possible. The survey contained 93 items and covered eight broad areas: family and geographical background, early military exposure, initial entry decision factors, previous employment, high school and college activities (both athletic and nonathletic), key life events, flying interest and experience, and initial military expectations and experiences. Most items took the form of biodata or background information. Where possible, questions referenced objective, verifiable events, in order to aid recall, increase accuracy, and facilitate implementation. Of a total of 15,264 survey packages mailed out, 9,409 (61.6%) were returned in usable form.

Study Design

Because of maturation differences, the criterion for success was defined differently for each cohort group. For the 1965-74 Cohort Group, "success" was defined as retention to 14 years of Total Active Federal Military Service (TAFMS), the point at which the officer attrition curve stabilizes. (Ninety-seven percent of officers who reach the 14-year mark stay to retirement). Success for the 1975-80 Cohort Group was defined as retention to 8 years (the point at which all officers have served their initial commitment, plus one year), since these officers have not yet had the opportunity to stay to 14 years. The 1981-84 Cohort Group was deleted from the present analyses but was reserved for future analyses as these officers mature to a minimum criterion of 8 years.

The Study Population Analyses used the full cohort population of 146,566 officers but were restricted to Study Population File data. These analyses were stratified by Cohort Group (1965-74 and 1975-80), Commissioning Source (USAFA, OTS Non-Prior Service, OTS Prior Service, and AFROTC), and Utilization Area (Pilot, Navigator, Engineer, and Other; 32 cells total). The Combined Analyses used both Study Population File data and Survey File data but were restricted to the survey sample of 9,409 officers. These analyses were stratified by Cohort Group, Commissioning Source, and Utilization Area (Pilots and Nonpilots only, due to the smaller number of subjects available on the survey file). Combined Analyses were omitted for the 1965-74 OTS Prior Service Pilots, due to the unexpectedly small number of subjects, 49, in this cell. This deletion left 15 cells in the Combined Analyses.

Results and Discussion

Two types of results are presented here: (a) effects of key predictor variables on retention and (b) results of the multiple linear regression modeling efforts.

Key Individual Predictors

Each variable below is identified as either a population (P) variable, based on the full cohort population, or a survey (S) variable, based on the survey sample. The .05 significance level is used for reporting findings in all cases. As a baseline, the overall retention levels were:

	1965-74 Group	1975-80 Group
Study Population	34.3%	67.8%
Survey Sample	35.6%	66.6%

Retention was significantly higher than average for:

1. (P) Navigators, followed by pilots, engineers, and other officers, respectively.
2. (P) OTS Prior Service graduates, followed by Academy graduates, AFROTC graduates, and OTS Non-Prior Service graduates, respectively. Among OTS graduates, Prior Service graduates with up to eight years prior enlisted time surpassed Non-Prior Service graduates.
3. (S) Officers whose parents were military members or blue-collar workers, as opposed to white-collar workers or self-employed.
4. (S) Officers whose parents were career enlisted members, as opposed to career officers.
5. (P) Graduates of moderate-cost ($5000 per year) colleges. Graduates of high-cost colleges ($7500 or more per year) had significantly lower retention.
6. (S) Officers whose parents/families had higher mobility.
7. (S) Officers who stated they joined the Air Force for job security.
8. (P) Academy graduates who were higher in the Military Order of Merit.
9. (P) AFROTC engineering graduates working in non-engineering specialties (as opposed to those working in engineering specialties.)

Retention was significantly lower than average for:
10. (S) Officers from higher-income, more highly educated families.
11. (S) Officers who stated they joined the Air Force for training.
12. (S) Officers who had a civilian pilot rating or large number of flying hours prior to entering their commissioning program.

Among factors that had little or no predictive value in retention (often contradicting commonly held assumptions) were:
14. (P) SAT and AFOQT scores.
15. (S) College GPA (although high school class standing correlated negatively for Nonpilots).
16. (S) High school and college activities (sports and non-sports).
17. (S) Scouting experience (including Eagle Scouts).
18. (S) Work history (including newspaper delivery experience).
19. (S) Geographic region (at time of high school graduation).
20. (S) Key life events (i.e., age at first date, traffic ticket, etc.).

Regression Models

Study Population Analyses. A separate regression model was constructed for each of the 16 cells (four commissioning sources X four utilization areas) for each cohort group. The Study Population File contained seventy-seven candidate variables. For the 1965-74 Cohort Group, R^2 values ranged from .006 to .023, with two cells failing to generate a model (i.e., no variables significantly predicted retention). For the 1975-80 Cohort Group, the R^2 values ranged from .009 to .034. These values are obviously quite low, at best accounting for some three percent of the variance in retention. These results indicate that the precommissioning data currently collected by the personnel system are insufficient to enhance current selection procedures and improve retention. Furthermore, they are insufficient to even accurately predict retention trends.

Combined Analyses. A separate regression model was constructed for each cell in the 1965-74 Cohort Group (7 cells) and 1975-80 Cohort Group (8 cells). For the 1965-74 Cohort Group, the R^2 values ranged from .193 to .660, with a median R^2 of .246. For the 1975-80 Cohort Group, the R^2 values ranged from .230 to .982, with a median R^2 of .422. (The .660 and .982 values are based on two OTS Prior Service cells, with high retention rates and relatively small cell sizes.)

The regression models were cross-validated by a jack-knifing procedure, in which each subject was sequentially removed from the sample, a new regression model constructed, and the model used to predict that subject's retention behavior. For the 1965-74 Cohort Group, cross-validated R^2 values ranged from .135 to .331, with a median R^2 of .177. For the 1975-80 Cohort Group, the R^2 values ranged from .122 to .723, with a median R^2 of .259. These cross-validated coefficients show some drop from the original values but a substantial amount of "true variance" captured by the original models, as well.

Conclusion

The Combined Analyses demonstrate that we can predict substantial variance in retention behavior from a precommissioning perspective. More important, the Air Force can use these variables to enhance current selection procedures by selecting officer candidates with a view to their likelihood of retention. This could be accomplished by including a "Retention Index" in each applicant's selection package. However, this increased prediction/ selection power cannot be obtained without substantially enhancing our current applicant information. ATC's Commissioning Programs Directorate is now examining both the utility and policy implications of this approach.

References

Cook, T.M., & Morrison, R.F (1983). Surface warfare junior officer retention: Background and first sea tour factors as predictors of continuance beyond obligated service (Report No. NPRDC TR 83-6). San Diego, CA: Navy Personnel Research and Development Center.

Dunn, J.T., & Feiler, J.P. (1983). A model of United States Air Force turnover (Report No. LSSR 99-83). Wright-Patterson AFB, OH: School of Systems and Logistics, Air Force Institute of Technology.

Finstuen, K., & Alley, W. E. (1983). Occupational and personnel correlates of first-term enlisted tenure in the Air Force (Report No. AFHRL-TR-82-36). Brooks AFB, TX: Manpower and Personnel Research Division, Air Force Human Resources Laboratory.

Gibb, G.D., Nontasak, T., & Doigin, D. L. (1988). Factors affecting career retention among aviators. Journal of Business and Psychology, 2 (4), 321-326.

Shenk, F., Watson, T.W., & Hazel, J.T. (1973). Relationship between personality traits and officer performance and retention criteria (Report No. AFHRL-TR-73-4). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Ward, M. P., & Tan, H. W. (1985). The retention of high-quality personnel in the U. S. armed forces (Report No. R-3117-MIL). Santa Monica, CA: RAND.

Watson, T. W., & Appel, V. H. (1986). Implications of the results of recent turnover research for Air Force policy (Report No. AFHRL-TP-86-5). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Author Notes

This paper is a condensation of the final report of the ATC Officer Selection Study Group, Officer selection: Can we select to retain?, available from HQ ATC/RSC, Randolph AFB, TX 78150. Questions on the current study should be addressed to the author. Questions on the data files and their applications should be addressed to Maj David K. Roberts, HQ ATC/RSCY.

The Officer Selection Study Group consisted of: Lt Col F. Randall Starbuck, Lt Col Merl A. Morehouse, Lt Col Daniel E. Hancock, Maj William H. Cummings, Maj David K. Roberts, Capt Glenwood M. Barbee, Capt Gary R. Grimes, SSgt Wayne K. Gatson, Sgt Phillip M. Harris, and Ms. Ingrid Wright.

The author would like to express his sincere thanks to Dr. Tom Watson and Ms. Jacobina Skinner for their comments on earlier drafts of this paper.

Performance Feedback:
Can an Expert System Assist Air Force Managers With This Task?

Lt Col Tim O. Peterson, Ph.D.
Air Force Logistics Management Center
David D. Van Fleet, Ph.D.
Arizona State University - West Campus

Abstract

This manuscript examines the possible application of expert systems technology to the managerial task of performance feedback. Proponents of expert systems argue that these knowledge rich systems can provide managers with assistance in accomplishing tasks such as performance feedback. In this manuscript, one such expert system, Performance Mentor, is described. In addition, empirical findings are presented using this expert system and Air Force managers.

In May 1987, the senior leadership of the Air Force embarked on a review of the officer corps' performance appraisal system (Dalton, May 18, 1987). In early 1988, a similar review of the enlisted performance appraisal system was launched (AFP 39-15, May 1, 1989). Today, the Air Force has two new performance appraisal systems, the Officer Evaluation System (OES) and the Enlisted Evaluation System (EES). The heart of both appraisal systems is the performance feedback session conducted by the immediate supervisor. These feedback sessions involve the supervisor sitting down with the employee for the sole purpose of providing work-related performance information.

Performance Feedback Process

Providing performance feedback to employees is considered one of the most important activities within an organization (Stoner & Wankel, 1986), and yet one of the most difficult for managers to perform (Ivancevich, Donnelly, & Gibson, 1986). Larson (1984) argues that feedback about an individual's performance is an integral component of any organizational control system. Without it, change in performance only occurs by accident. Unfortunately, supervisors seem to resist giving performance feedback, particularly if the feedback is negative in nature (Fisher, 1979). However, Hawkins, Penley, and Peterson (1981) found that employees' performance ratings were higher for those employees who reported receiving high levels of performance communication from their supervisors. Thus, it is imperative that organizations like the Air Force ensure that managers not only provide performance feedback, but that those managers do it well.

Managers find it difficult to provide performance feedback to their subordinates for two primary reasons. The first is that managers realize they do not possess the necessary skills (knowledge, cognitions, and behaviors) to effectively perform the task of giving negative feedback (Carroll, 1982; Lefton, 1985). Because they do not possess these skills, managers realize that performing this task ineffectively could damage or destroy the interpersonal relationships they have developed with their subordinates. Second, having come to this conclusion about their skill deficiencies, managers develop low self-efficacy expectations (Bandura, 1977; Bandura &

Schunk, 1981) about their ability to successfully provide negative feedback while maintaining a positive interpersonal relationship with that subordinate.

Since Maier (1958) wrote that the reason for the failure to provide performance feedback was attributable to a lack of managerial skill, organizational practitioners and management scholars alike have attempted three different ways to provide inexperienced managers with the necessary knowledge base to effectively conduct feedback sessions. These methods (social learning, training, and education) have not been particularly successful.

A potential solution to this problem could be the use of expert systems. Expert systems that claim to provide managers with the necessary knowledge and skills to perform such managerial functions as planning, organizing, directing, and controlling are currently being developed (Silverman, 1987). More recently, expert systems that claim to provide managers with the necessary knowledge and skills to perform managerial tasks such as providing performance feedback have appeared (Blanning, 1987; Kearsley, 1987). One of those, Performance Mentor, seems particularly useful to Air Force managers.

Performance Mentor

Performance Mentor is a commercially available rule-based expert system with over 350 rules. According to the developers, the advice provided by the system is based on a massive literature search by the knowledge engineer and utilizes material from over 100 sources (Schlitz, 1986). Performance Mentor is designed to assist the manager in the feedback process, not in the evaluation. The first step for a new user of Performance Mentor is to profile himself/herself and the organization. The managerial profile is done by selecting descriptive terms (e.g. relaxed, charming, precise, cautious) from 59 different terms offered by the software. Similarly, the manager selects from 45 terms (e.g isolated, pragmatic, free, friendly) to profile the workplace. To complete the workplace profile and managerial style profile, 38 multiple choice questions are asked. They deal with such things as the user's experience giving performance appraisals. This information is stored by the system in a separate database so that it can be updated when changes occur.

Next, the software asks the manager to profile each of his/her subordinates. Using 68 terms similar to those used for the manager, each subordinate would be described. After completing these profiles, the manager can obtain advice from Performance Mentor. During an advisory session, the system asks a series of 66 multiple choice questions. These deal with such matters as the particular person's experience on the job and his/her responses to previous performance feedback sessions. Performance Mentor also asks the manager to describe the immediate work setting and the type of feedback which will be provided to the subordinate. The system then uses all of this information to recommend a performance feedback interviewing strategy for each superior-subordinate dyad. These reports typically run two to four pages. While reviewing the advice on the screen, it is possible to get explanations as to why that advice is being given. Those explanations frequently serve to teach the user the "why" behind recommendations.

Performance Mentor takes a broad approach to performance feedback. It incorporates empirical findings from the feedback literature as well as from the literature about human relations, interpersonal communications, and personality. For these reasons, Performance Mentor seems to provide a complete and realistic recommendation on how to conduct performance feedback interviews.

While the claims of proponents of expert systems such as Performance Mentor seem promising, those claims have not been supported by carefully collected, objective research data as opposed to more subjective, anecdotal reports from users. As Dr. Arie Lewin (1986) said at the 1986 Annual Academy of Management Conference, "Whether these new systems help managers is still a question to be answered. We just don't know. At present, we are working on faith."

Empirical Finding

Peterson (1988) designed an experiment to test this question using Performance Mentor. The subjects (N=60) were all Air Force managers. Half of the managers used the expert system while the other half did not. The subjects performed the role of a supervisor of three purchasing agents. The setting was a highly realistic task involving extensive paper documentation of performance as well as a 15 minute long videotape of the individual to be evaluated showing him at work.

The important point was not the actual performance appraisal evaluation (which was designed to be negative), but the feedback session. In this study, the focus was on the managers ability to identify the appropriate behaviors for a negative performance feedback session. The subjects had to specify what they felt where the ten correct behaviors from a list of 50 possible behavioral responses which included both correct things to do such as "acknowledge his experience and tell him you know he can do the job if he wants to" and incorrect things to do such as "demand an explanation for his sloppy ordering over the last six months." In addition, the list of possible behavioral responses also included some neutral behavioral responses such as "call the individual in immediately." These responses are not intended for sharing information or feelings; they are used only to open the lines of communication for more meaningful communication later on (Verderber, 1981).

The results clearly support the hypothesis that an expert systems can really help Air Force managers do their jobs better. Managers using the expert system performed 400 percent better on the task of identifying correct behavioral responses than the managers not using the expert system. Further support for the hypothesis can be gleaned by examining the number of incorrect behaviors chosen by the two groups. Managers using the expert system identified fewer of the incorrect behaviors as being appropriate than did the managers not using the system (6% versus 17%). These findings are significantly different at the .05 level using the student's t-test statistic. Whether these managers would follow through on their intentions and effectively apply that knowledge is another issue. But, these results suggest that the potential is there since the expert system did aid the mangers in identifying both the proper behavioral responses and the improper behavioral responses.

One aspect which was not directly tested or analyzed, but which was observed, was that the use of the expert system, while substantially improving the quality of the process, seemed to slow down the decision process. On the average, it took managers using the expert system 15 minutes more to come to a decision than it did the managers not using the expert system. From a cost/benefit perspective, the use of the expert system still appeared worthwhile considering the increase in appropriate behavioral responses achieved over this period of time. This type of information will be essential as one evaluates the costs and benefits of the application of expert systems to managerial tasks.

Implications

The results above suggest a number of important implications for managers and the Air Force. First, while managers are not generally trained in providing performance feedback, the training involved in this study was extremely brief, and yet it yielded good results. If similar brief training sessions yield similar results with other expert systems, then, training time and training costs can be reduced using expert systems. This would mean that new Air Force managers could more rapidly reach the levels of effectiveness of experienced managers.

Another implication is that the use of an expert system will slow the decision making process, but will provide a better solution to the specific performance feedback situation. While most computer technology is billed as providing information faster, expert systems may actually slow the process. In this study, the expert system slowed the manager's decision-making by about fifteen minutes, but increased their performance feedback knowledge fourfold. This slowing of the decision process seems to be beneficial in that it allows more alternatives to be examined than would be examined without the system. This certainly seems like a reasonable tradeoff given the importance of providing performance feedback to subordinates and the difficulty that managers have with this managerial task.

A final implication for organizations is that if expert systems can provide expert advice on one managerial task, they may be able to provide advice on many of the other managerial tasks that face managers. The results of this study show that expert systems can do these things quite impressively. The challenge of the future for organizations will be extracting and organizing the knowledge they have regarding the various managerial tasks to be used in building expert systems. The knowledge and technology to do this are now available.

Conclusions

Clearly, then, managers and organizations may benefit from the use of expert systems even for areas of administration which are not routine and quantifiable. Furthermore, the authors believe that this study and future research into managerial expert systems will provide management scholars and practicing managers with the knowledge and skills necessary to perform complex managerial tasks such as providing performance feedback in a confident, professional, and humane manner. By so doing, organizations will experience an increase in human productivity which is ultimately what organizations pay managers to achieve.

References

Air Force Pamphlet 39-15. (May 1, 1989). USAF Guide to the Airman and NCO Performance Feedback System. Washington DC: Government Printing Office.

Bandura, A. (1977) Self-efficacy: Toward a unifying theory of behavioral change. Psychological Review, 84, 191-215.

Bandura, A. & Schunk, D. H. (1981) Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. Journal of Personality and Social Psychology, 41, 586-598.

Blanning, R. W. (1987) A survey of issues in expert systems for management. In B. G. Silverman (Ed.), Expert systems for business. Reading, MA: Addison-Wesley, 24-39.

Carroll, S. J. (1982) The performance review and quality of work life programs. Paper presented at the Annual Meeting of the Eastern Academy of Management, Baltimore, Maryland.

Dalton, P. (1978, 18 May). 3 groups study new OER, but goal is "Get It Right". Air Force Times, 47, 6.

Fisher, C. D. (1979) Transmission of positive and negative feedback to subordinates: A laboratory investigation. Journal of Applied Psychology, 64, 533-540.

Hawkins, B. L, Penley, L. L., & Peterson, T. O. (1981) Developing - Behaviorally Oriented Communication Scales for the Study of the Communication - Performance Relationship, Paper presented at the annual meeting of the Academy of Management, San Diego.

Ivancevich, J. M., Donnelly, J. H., Jr., & Gibson, J. L. (1986). Managing for performance (3rd Edition). Dallas, TX: Business Publications, Inc.

Kearsley, G. (1987) Software for the management of training. Training News, 8, 13-15.

Larson, J. R., Jr. (1984) The performance feedback process: A preliminary model. Organizational Behavior and Human Performance, 33, 42-76.

Lefton, R. E. (1985) Performance appraisal: Why they go wrong and how to do them right. National Productivity Review, 5, 54-63.

Lewin, A. Y. (1986) Presentation at the annual Academy of Management Conference. Chicago, Illinois.

Maier, N. R. F. (1958) The appraisal interview: Objectives, methods, and skills. London, UK: Wiley.

Peterson, T.O. (1988). The acquisition of managerial performance feedback skills through the use of a knowledge-based expert system: An empirical evaluation. Unpublished Dissertation. Texas A&M University, College Station, TX.

Schlitz, J. (1986) US Government and AI Mentor Inc. Sign Site License Agreement. Press Release, AI Mentor, Inc., October 15, 1986.

Silverman, B. G. (1987) Should a manager 'hire' an expert system? In B. G. Silverman (Ed.), Expert Systems for Business. Reading, MA: Addison-Wesley, 5-23.

Verderber, R. F. (1981). Communicate! (3rd Edition). Belmont, CA: Wadsworth Publishing Company.

Factors Affecting Career Decisions of Air Force Female Aviators

Second Lieutenant Beth L. Schissel
United States Air Force Academy

Abstract

This article examines the career decision factors of Air Force female aviators. The major decision factors differ depending on the cited career intention. Family concerns, due in large part to the current joint spouse policy, are the primary decision factor for all female aviators. Job satisfaction concerns are minor in comparison to overall career concerns. Female aviators cite limited command opportunity and the effects of current combat exclusion laws and policies as major career decision factors. The current policies regarding combat exclusion and joint spouse are seen as structural barriers to career advancement.

After a "ruling by the Canadian Human Rights Tribunal that combat exclusion was discriminatory," the ensuing integration of women into combat roles in the Canadian forces began in March of 1989 (Bird, 1989, p. 16). The events in Canada and the more recent United States invasion of Panama have resurfaced the debate over the validity of current United States laws and policies excluding women from combat roles.

A debate aired on *Face the Nation* barely a week after the United States began withdrawal of troops from Panama (Stahl, 1990). Representative Pat Schroeder, who favors allowing women in combat, debated the validity of reasons cited by Brian Mitchell, author of The Weak Link, for keeping women out of combat roles. Congresswoman Schroeder recommended a four-year trial of integrating women into combat roles, similar to the trial that preceded the full role integration of women into the Canadian forces.

Currently, the Air Force continues to redefine combat aircraft and missions. These efforts have resulted in opening 98 percent of all career fields and 96 percent of all positions to women (Bird, 1989). Unfortunately the 4 percent of the restricted positions include some of the most coveted jobs in the service, including flying bomber and fighter aircraft. Those positions, in fighters and bombers, are also part of the traditional path to key command positions and high-level promotions.

Previous research into the combat exclusion issue focused upon the opinion of female pilots in the United States Air Force (USAF) as to whether they felt they could perform the combat mission (Peterson, 1988). The conclusions of this study showed that USAF women pilots not only felt that they had the ability to fly combat missions, but that not being allowed to fly combat aircraft hindered their career and halted upward progression.

Further research into military women's perceptions (e.g., integration issues and the opportunity for advancement) was conducted in a *Government Executive* survey which focused on the opinions of senior women line officers (0-5, 0-6 and 0-7) within the Department of Defense (Sherwood, 1989). The results of this survey indicated that women feel that they are not advancing fast enough in their military careers, and they cited both structural barriers (presented by the combat exclusion laws and policies) and cultural barriers (e.g., sexual harassment and interaction with superiors).

Previous research suggests that women face both cultural and structural barriers as they progress in the military. In light of recent events regarding women in combat, this study solicits the perceptions of those women most affected by these barriers-- Air Force female aviators. What do these women consider as factors affecting their future career decisions?

Method

The descriptive data reported here are a small part of a larger study of female Air Force aviators who are members of Women Military Aviators, Inc. (WMA). The survey respondents, pilots and navigators, completed a 140-item questionnaire that included various demographic questions and both cultural and structural attitudinal questions. The questionnaire, sent out in early November, drew responses from 120 active duty Air Force female aviators (a 53 percent return rate). These officers represent approximately 22 percent of the total population of Air Force female aviators. An ideal study would include all Air Force female aviators, but time and cost constraints prohibited a more comprehensive survey. However, there is no reason to believe that this sample does not adequately represent the views of all Air Force female aviators. Demographically, the survey respondents are diverse in terms of age, years of service, military rank, and major command affiliation.

Results

Table 1 presents a comparison of female aviators who indicated their future career intentions (i.e., remain, separate, or undecided). These three samples of women aviators were analyzed by career decision factors (i.e., family, job, career, and other).

Table 1

CAREER DECISION FACTOR

		N	Family	Job	Career	Other*
	Remain in AF	36	41.7%	13.9%	33.3%	11.1%
CAREER INTENTION	Separate	36	58.3%	11.1%	25.0%	5.6%
	Undecided	37	45.9%	24.3%	16.2%	13.5%

*Content analysis of written statements showed no distinct pattern of concerns

Regardless of career intentions, family concerns are most often cited as a factor for future career decisions. When asked to rate their number one family concern, 44 percent of the respondents indicated that the joint spouse assignment policy was a major dissatisfaction. However, family concerns are lessened for those women who plan to remain in the Air Force.

The second greatest concern was reported to be career factors which increase for those who plan to make the Air Force a career. Thirty-six percent of the women's career concerns centered around structural barriers imposed by combat exclusion laws (i.e., promotion opportunity, limited command opportunity and combat restrictions). While not a majority, the remaining concerns were divided amongst eight diverse categories.

The third factor dealt with job concerns (e.g., lack of responsibility and extra military duties beyond flying). Although this category was rated relatively low in comparison to the others, 30 percent of those surveyed listed assignment location as their primary job concern. For those undecided, there is a shift from family and career concerns to concern about job satisfaction.

Discussion

The results of the study seem to indicate a relationship between career intentions and career decision factors. Regardless of indicated career intention, family concerns play a major role in career decision for Air Force female aviators. Approximately one-third plan to separate from the Air Force, citing family concerns as the reason. The remaining two-thirds of the women (i.e., undecided and plan to stay) also report that the current joint spouse policy is a major factor in their future career decision.

These women who are still undecided about their career cite family concerns as a primary career decision factor; however, they also are concerned with job satisfaction. Since they are undecided about their future career, their focus is on daily job satisfaction-- is this a job upon which to develop a career?

For women who plan to make the Air Force a career, family concerns are still paramount. These women have accepted some type of balance, ranging from complete sacrifice of familial roles to an assumption of multiple roles as worker, wife, and mother. Now, an increased factor becomes career concerns. They like their jobs, but are concerned about career progression, specifically related to the issue of combat roles. Eighty-five percent cited limited command opportunites and combat exclusion laws as deterrents to career progression. In comparison to the *Government Executive* survey (in which 55 percent of senior female officers reported that women are not advancing fast enough), 75 percent of the surveyed female aviators feel women are not advancing fast enough. Their more negative disparity in opinion may reflect how they are affected daily by the current combat exclusion laws and policies-- they can't fly fighter aircraft which often lead to key command positions.

Conclusion

Today, with the retention problems among aviators, Air Force leaders must seek solutions to retain flyers. To retain female aviators, structural barriers must be eliminated. Present combat exclusion laws and policies are major career decision factors. These laws and policies need to be rescinded in order to retain female aviators.

While combat exclusion laws affect primarily women, joint spouse issues are a potential concern for both men and women aviators; it is not gender specific. The potential for the current joint spouse policy to have a negative bearing upon career decisions of aviators can only add to the current manpower crisis. It is time to implement policies to retain our highly trained aviators, both men and women.

References

Bird, J. (1989, December 4). Breaking the barrier: In canada women now fill combat jobs. Air Force Times, pp. 16-17,76.

Peterson, T. (1988). USAF women pilots: The combat issue. Air Command and Staff College Student Report #88-2110. Maxwell AFB, AL: Air University.

Sherwood, D. (1989). Women in the military. Government Executive, 8, 10-19,56.

Stahl, L. (1990, January 14). [Interview with Representative Pat Schroeder and Brian Mitchell]. Face the Nation. CBS.

Author's Note

Views expressed in this paper are those of the author and do not necessarily reflect those of the United States Air Force Academy or any other government agency.

Person-Situation Effects in the Prediction of Organizational Commitment[1]

L. Alan Witt, Ph.D.
Western Illinois University
LtCol Mickey R. Dansby, Ph.D.
Defense Equal Opportunity Management Institute

Abstract

Recent research (Eisenberger, Huntington, Hutchinson, & Sowa, 1986; Kottke & Sharafinski, 1988) suggests that the individual's perception of organizational support forms a significant component of the individual's commitment to the organization. Following recent developments in affect research, the present study examined the interaction between dispositional positive affect and organizational support in the prediction of organizational commitment. Data collected from personnel working in a Department of Defense training facility indicated that positive affect moderated the relationship between perceptions of organizational support and organizational commitment. These results have implications for research on organizational commitment.

For some time now, organizational researchers have studied the construct of organizational commitment and its relationships with various situational characteristics, attitudes, and behaviors. Commitment is an issue of considerable importance to the Department of Defense. Indeed, research has been based on the explicit assumptions that commitment is a potential determinant of performance (DeCotiis & Summers, 1987), turnover and absenteeism (Zaccaro & Collins, 1988), and prosocial behavior (O'Reilly & Chatman, 1986) and that the primary antecedents of commitment are within command's ability to influence (Angle & Perry, 1983).

The objective of the present study was to further explore the conceptual bounds of organizational commitment. Adopting an interactional psychology perspective, the introduction draws parallels between research on organizational commitment and related concepts at the situation and person levels, namely organizational support and dispositional positive affect.

While studies of organizational commitment have examined both person and situational variables, they have seldom examined the interaction between person and situational antecedents of commitment. Proponents of the interactional psychology perspective assert that behavior is a function of an interaction between the person and situation. The person is seen as an active participant in the interaction process, changing the situation and being changed by it (Terborg, 1981). This perspective promotes explicit assessment of situational variance in cues, rewards, or opportunities and individual variance in cognitions, abilities, or motivation. However, an important element on the person side of the interaction has been overlooked: dispositional positive affect.

[1]This paper reflects the opinion of the authors only and should not be construed as official policy of the Department of Defense or any organization.

To predict organizational commitment, we examined situational variance in rewards and opportunities operationalized as the individual's perceptions of support given by the organization. Individual variance in affect was assessed in terms of the individual's dispositional positive affect.

A considerable literature has examined the situational antecedents of commitment. Recent research (Eisenberger, Huntington, Hutchinson, & Sowa, 1986; Kottke & Sharafinski, 1988) suggests that the individual's perceptions of organizational support have a significant bearing on commitment to the organization. Eisenberger et al. (1986) argued that individuals who consistently receive formal support from the organization (e.g., provision of necessary equipment and time to get work done, organization leaders "standing up" for individuals under fire) and informal support from their colleagues (e.g., assistance with a heavy work load, emotional help in times of stress) are more likely to be committed to the organization.

Attempts to identify person-level influences of commitment have focused on such factors as demographic variables, attributional tendencies, and stress. We suggest that an individual component of commitment to the organization is the individual's dispositional positive affect.

Recent developments in affect theory and measurement indicate that positive and negative affect are independent dimensions of disposition (Diener & Emmons, 1985). Positive affect has been defined as "reflecting the extent to which a person feels enthusiastic, active, and alert" (Watson, Clark, & Tellegen, 1988, p. 1063). Individuals high in positive affect tend to experience high energy, full concentration, and pleasurable engagement, whereas individuals low in positive affect tend to experience lethargy and sadness.

In line with the definition of positive affect, we hypothesized that positive affect scores would be positively correlated with perceptions of organizational support and commitment. Individuals who are dispositionally more enthusiastic and energetic may be more likely to perceive greater support from the organization and manifest commitment to it.

In line with the interactional psychology perspective, we expected perceptions of support to differentially account for variance in organizational commitment among individuals with different levels of dispositional positive affect.

Since individuals low in positive affect tend to be sadder and more lethargic, it is likely that they would look more to the work environment in the development of their commitment; because individuals high in positive affect tend to be dispositionally enthusiastic, the work environment would be less salient in the development of their commitment. Thus, perceptions of organizational support should be more highly correlated with organizational commitment among individuals low in positive affect than among individuals high in positive affect.

Method

Forty-nine (59%) of the 83 available personnel working in a DoD training facility in the eastern U.S. voluntarily and anonymously completed an organizational survey including the Eisenberger et al. (1986) measure of perceived organizational support ($\bar{X}$ = 47.17, SD = 6.9, alpha = .94), the Watson, Clark, & Tellegen (1988) positive affect scale ($\bar{X}$ = 35.37; SD = 7.4; alpha = .97), and the Hrebiniak and Alutto (1972) measure of organizational commitment ($\bar{X}$ = 12.07; SD = 4.4; alpha = .79). High scores on the support scale indicate perceptions of greater support. High scores on the positive affect scale indicate higher positive affect. High scores on the commitment scale indicate greater commitment to the organization. Directions for the positive affect scale instruct individuals to indicate how they generally feel by responding on a 5-point scale (1 = very slightly or not at all; 5 = extremely) to 10 adjectives (e.g., enthusiastic); this scale assesses how individuals feel generally (i.e., trait or dispositional positive affect) rather than state positive affect (Watson & Clark, 1984).

Results and Discussion

Positive affect scores were positively related to both support (r = .50, p < .01) and commitment scores (r = .52, p < .01). Support and commitment scores were positively related (r = .83, p < .01).

Moderated multiple regression was used to assess the moderating effect of positive affect by adding the cross-product term as a separate predictor in the equation (Zedeck, 1971). The commitment scale scores were regressed on positive affect scale scores, support scale scores, and their cross-product term. Hierarchical multiple regression was then used with the positive affect scores and support scores entered first and their cross-product term entered second. The significance of the incremental R^2 (ΔR^2) with the addition of the cross-product term was assessed. The addition of the cross-product term added significant variance to the prediction of commitment (full model adjusted R^2 = .388410; reduced model adjusted R^2 = .36626; ΔR^2 = .01784, F = 24.95, p < .01).

In order to assess the direction of the interaction and to test the hypothesis that the support-commitment relationship would be greater among individuals low in positive affect than those high in positive affect, the subgrouping method of examining differential correlations was employed (cf. Arnold, 1982). Scores on the positive affect scale were split on the median, creating groups low and high in positive affect. Zero-order correlations between the scores on the organizational support scale and organizational commitment scales were then computed for both the low and high groups. Confirming the hypothesis, the results indicated that support was more highly related to commitment among individuals low in positive affect (r = .83, p < .01) than among individuals high in positive affect (r = .21, ns). Because this procedure is subject to a number of problems, this was not the primary method used to assess the interaction. However, it permitted assessment of the comparative strength of the correlations between the low and high positive affect groups.

Discussion

Congruent with the hypotheses and the interactional psychology perspective, the results suggest that perceptions of support differentially account for variance in organizational commitment among individuals with different levels of dispositional positive affect.

Limitations of the Study

Several limitations of the present study should be noted before implications are discussed. First, data were collected from a small sample of personnel working in a DoD training organization, which may not be representative of most DoD organizations.

Second, a number of variables that may have accounted for the findings reported here or additional variance were not assessed. These include personnel multiple commitments, personnel rank, conflict between family and job roles, and other person-level issues.

Third, because it may also be suggested that individuals who experience certain levels of support and commitment may report certain dispositions, it is possible that support may be the moderator of the positive affect-commitment relationship rather than the other way around.

Fourth, because of organizational considerations, most of the participants completing the questionnaires did so in one sitting. Thus, these data may be subject to common method variance. Indeed, in indicating the moderating effect of dispositional positive affect on the support-commitment relationship, the data provide some support for concerns about method variance (i.e., by demonstrating the effect of disposition on the relationship between perceptions of the work environment and commitment).

Implications for Management Practice

Efforts to increase support from the organization (or change perceptions of what is being provided) in order to promote commitment may be more successful among individuals low in positive affect than among individuals high in positive affect. Thus, organization development interventions might take into account individual differences in dispositional positive affect in order to enhance their effectiveness.

Implications for Theory and Research

The data reported here support the interactional approach to the study of organizational commitment. Researchers have shown that commitment is multidimensional. The interactional approach may be useful for identifying antecedents of this multidimensionality. Moreover, additional variance in organizational behaviors may be explained by using this approach.

References

Angle, H. L., & Perry, J. L. (1983). Organizational commitment: Individual and organizational influences. *Work and Occupations, 10*, 123-146.

Arnold, H. J. (1982). Moderator variables: A clarification of conceptual, analytic, and psychometric issues. *Organizational Behavior and Human Performance, 29*, 143-174.

DeCotiis, T. A., & Summers, T. P. (1987). A path analysis of a model of the antecedents and consequences of organizational commitment. *Human Relations, 40*, 445-470.

Diener, E., & Emmons, R. A. (1985). The independence of positive and negative affect. Journal of Personality and Social Psychology, 47, 1105-1117.

Eisenberger, R., Huntington, R., Hutchinson, S., & Sowa, D. (1986). Perceived organizational support. Journal of Applied Psychology, 71, 500-507.

Hrebiniak, L. G., & Alutto, J. A. (1972). Personal and role related factors in the development of organizational commitment. Administrative Science Quarterly, 17, 555-572.

Kottke, J. L. & Sharafinski, C. E. (1988). Measuring perceived supervisory and organizational support. Educational and Psychological Measurement, 48, 1075-1079.

O'Reilly, C. A., & Chatman, J. (1986). Organizational commitment and psychological attachment: The effects of compliance, identification, and internalization on prosocial behavior. Journal of Applied Psychology, 71, 492-499.

Terborg, J. R. (1981). Interactional psychology and research on human behavior in organizations. Academy of Management Review, 6, 569-576.

Watson, D., & Clark, L. A. (1984). Negative affectivity: The disposition to experience aversive emotional states. Psychological Bulletin, 96, 465-490.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. Journal of Personality and Social Psychology, 54, 1063-1070.

Zaccaro, S. J., & Collins, T. T. (1988). Excused and unexcused absenteeism in normative organizations: Effects of organizational commitment, rank, and interaction process. Group and Organization Studies, 13, 81-99.

Zedeck, S. (1971). Problems with the use of "moderator" variables. Psychological Bulletin, 76, 295-310.

Who says the Recruiter's Job is Stressful?[1]

Herbert George Baker, PhD
Navy Personnel Research and Development Center

Abstract

This paper brings together evidence from a variety of sources to support the contention that the job of the Navy recruiter is pervaded by stress. Job stress is caused by factors ranging from recruiter selection and training, through incentives systems, to organizational policies and mangement styles.

By dint of great effort, the Navy recruiting force usually has been able to achieve its goals relative to both the quantity of accessions and quality of recruits. However, this has not been without its costs. One cost has been a decrement in recruiter quality of life. Job stress has become a pervasive problem in Navy recruiting, affecting personnel at all levels of the organization. While stress has many positive aspects, excessive job stress or an inept response to it is inimical to quality of personal and family life, individual and work unit performance, and organizational productivity.

Evidence of Stress

The MEDCOM Survey

In 1988, the Naval Medical Command (MEDCOM) conducted a brief informal survey of some 200 health practitioners eliciting information on both Navy and Marine Corps recruiters. The survey results provide evidence that stress and certain behaviors which are often positively correlated with stress are problems among recruiters. The Navy medical community is involved more than occasionally in personnel evaluation and/or treatment.

More than half (58.1%) of the respondents reported having "seen a Navy or Marine Corps recruiter for stress or other psychological problems." In all, 309 recruiters had been evaluated for continuance in their jobs. Two hundred and seven recruiters had received treatment. Of those who had treated a recruiter, 21.9 percent also had treated a spouse or other member of the recruiter's family. The most common diagnoses are listed in Table 1.

Table 1. Most Common Diagnoses of Recruiters Seen for Stress or Other Psychological Problems

Marital Dysfunction	34.1%
Stress/Depression	31.8%
Adjustment Disorder	15.9%
Alcohol and Drug Abuse	15.9%
Mixed Emotional Feelings	2.3%

Of those respondents who had treated a recruiter and/or spouse, 14.4 percent reported having included child abuse and 18.8 percent reported having included spouse abuse in the diagnoses. Furthermore, a number of abusive or addictive behaviors often associated with stress were included in the diagnoses. These are reflected in Table 2.

[1]Views and opinions are those of the author and do not necessarily represent those of the Navy or the Department of Defense.

Table 2. Abusive or Addictive Behaviors Noted

Alcohol	38.8%
Drug	12.2%
Food	8.8%
Gambling	6.6%

The consensus was: "Apparently, all psychologists agree that the stress recruiters face/experience is far greater/much more intense than shipboard environment". COMNAVCRUITCOM stated his intention to pursue stress training "for all of Navy Recruiting Command and their spouses".

Naval Inspector General's (IG) Summary Report to COMNAVCRUITCOM

In 1988, the IG conducted an inspection of Headquarters, CRUITCOM, and "noted with concern the overriding emphasis within NAVCRUITCOM on meeting goal every 30 days which often resulted in inadequate command focus on human requirements of individual recruiters and their families." The IG's report also spoke to: (1) the burdens borne by families, many of which had to live in areas which were high cost and/or remote from military support groups and facilities, and (2) the high stress of the work environment as contributor to a quality of life among Navy recruiters that was significantly lower than that of "their peers assigned to other types of shore duty." The IG's report described the work climate as "primarily one of extreme personal pressure, long working hours, and remoteness from Military Support Systems." The IG cited the following as factors negatively affecting the work environment: (1) constant pressure month after month to achieve goal, (2) long working hours with leave and liberty as uncertainties, (3) intolerance of failure to achieve production goals, (4) lack of support from supervisors, (5) reduction of tools (i.e., advertising, etc.) from the Navy to support "sales" efforts, and (6) reduced resources, occasionally forcing a cap on some operating expenses.

No-Fault Transfers

Under the incentive and tracking system for recruiters, individuals who, through no fault of their own, cannot become successful (i.e., productive) recruiters may be transferred out of recruiting. This is called a "no-fault" transfer, which is not supposed to exert a negative career impact. Typical reasons for no-fault transfers are: medical, lack of salesmanship, and lack of fluency in spoken English. Stress and stress-related problems are also valid reasons. As evidenced by the No-Fault Transfer Log maintained by NAVCRUITCOM, there were 45 no-fault transfers during the period 1 October 1988 through 30 June 1989. Of these, many were for conditions often considered stress-related; 14 of them actually used the word "stress" in the "reason for transfer."

Interviews with Navy Recruiters and Recruiting Managers

In February 1989, researcher teams from the Navy Personnel Research and Development Center (NAVPERSRANDCEN) interviewed 150 recruiters, plus a number of supervisorial personnel in the field and managerial personnel at several Navy Recruiting District (NRD) headquarters (Robertson, 1989). In all, interviews were conducted with recruiters at 29 Recruiting Stations and with managers at 11 NRDs.

Stress and pressures of the job and their adverse effects on personal and family life were noted over and over again in the comments made to the interviewers. It was apparent to the interviewers that stress was endemic among the Navy recruiting personnel. In some cases it led to difficulty:

> Some of the ensuing stress-related ailments are severe. In one district which has usually been at or near the top of making goals during recent years,

> casualties included two heart attacks, two referred for alcohol rehabilitation, two referred for being overweight, four referred for counseling, and two cases of family problems. One researcher observed that several recruiters described symptoms resembling those of combat fatigue stress disorders–dreaming about numbers and talking or yelling in their sleep. (Robertson, 1989, p. 9)

Relevant conclusions from the interviews include: (1) selection processes permit too many individuals who are not suited to the job to be assigned to recruiting duty, (2) potent stressors include extremely long work hours due to the pressure to meet monthly goals and having to deal with uncertainties, (3) stress management and related training should be implemented at the recruiter school and in refresher training courses, (4) stress may also be caused by onerous paperwork requirements, and (5) some job stress has its origins in the incentive and awards system and in the no-fault transfers that are incorporated within it.

Based on the information contained in the interviews, Robertson (1989) developed a work stress model that posits a "chain reaction" whereby intense pressure to make goal, plus a number of elements outside the recruiter's control, lead to: (1) a sense of failure, (2) stress-related ailments, and (3) reduced productivity.

The 1989 Recruiter Survey

A survey of all production recruiters (n=3,498; 3,252 responses used in analyses) was conducted by NAVPERSRANDCEN at the request of the Recruiting Long-Range Plan Study Group (RLRPSG), which had undertaken to examine all aspects of Navy recruiting. The researchers found additional confirmation that the job of the Navy recruiter is a stressful one, and that a multitude of potential stressors pervade the recruiting occupational milieu (Baker, Somer, & Murphy, 1989).

Tables 3 and 4 reflect with the number of hours worked per week and the number of days leave taken in the past year. These two factors make for continued and largely unrelieved pressures on the job, leaving little time for family or friends, for study for advancement in rating, or for taking care of personal business. And, given the 30-day goaling cycle, there is no letup–finish one month and begin immediately on the next. For the recruiter who avoids transfer, this translates into 3 years of intense, sustained pressure; or, as one recruiter termed the situation, "Thirty-six one-month tours."

Recruiter responses to the survey provide evidence of potential contributors to occupational stress in almost every area of the job: (1) people unsuited for sales work, or people with personal or family problems that can only be exacerbated by recruiting duty being selected, (2) previews of the job are inadequate resulting in erroneous expectations, (3) insufficient training time and resources are directed to stress awareness and stress management at the recruiter school and on the job, (4) performance management systems (supervisory techniques, goaling methods, performance monitoring and incentive plans, etc.) are being used that increase pressure on the recruiter but do not appear to the recruiter to support his or her job, (5) provisions for individual career development for recruiters are inadequate, and (6) resources to support the work of the field recruiter are too low.

The recruiter survey also provided an opportunity for recruiters to respond freely about their duties and their quality of life. A majority of the survey respondents (60.2%, n = 1,996) completed the free-response write-in section. Of those submitted, 1,935 (97%) were subjected to a content analysis (Aunins, Sander, Giannetto, & Wilson, in press).

Table 3. Hours Worked in a Typical Work Week
(from RLRPSG, 1989)

C18 Question: How many hours do you work in a typical work week?

	Frequency	Percent	Cum Percent
40 through 49	71	2.2	2.3
50 through 59	384	11.8	14.7
60 through 69	1,177	36.2	52.7
70 through 79	1,033	31.8	86.1
80 through 89	323	9.9	96.5
90 and above	107	3.3	100.0
No response	157	4.8	
Total	3,252	100.0	

70 hours or more 47%
60 hours or more 85%

Table 4. Days of Leave Taken During the Past Year
(from RLRPSG, 1989)

C19 Question: How many days of leave did you take during the past year?

	Frequency	Percent	Cum Percent
No leave taken	735	22.6	22.9
1-5 days	397	12.2	35.2
6-10 days	595	18.3	53.7
11-15 days	648	19.9	73.9
16-20 days	258	7.9	81.9
More than 20 days	582	17.9	100.0
No response	37	1.1	
Total	3,252	99.9	

Stress ranked seventh in a list of the 10 most common dissatisfiers. Three hundred sixteen recruiters made negative comments about psychological stress, and 50 about physiological stress. Comments indicating the presence of family conflict were made by 357 respondents. Ten recruiters mentioned suicide. Other contributors to job stress for the recruiter include: goaling pressures, excessive working hours, and inept or unresponsive supervision. Negative comments about working hours ranked number one in the list of the 10 most common dissatisfiers, with global comments about goals ranking third, and comments about "hammer management" by supervisors coming in fifth.

The presence within recruiting of so large a number of known stressors compellingly substantiates the assertion that recruiting is a high stress job. The responses to three items from the objective portion of the recruiter survey graphically communicate this message:

- Over 89 percent agreed with the statement, "My job as a recruiter is more stressful relative to other Navy jobs."

- More than 81 percent agreed with the statement, "I feel much stress in my job."
- Sixty-two percent agreed with the statement, "Job stress is a problem for me."

Discussion

Clearly, the evidence points to a high, perhaps dangerously high, level of stress extant throughout NAVCRUITCOM, and more than adequate supports the assertion that stress is a pervasive problem for NAVCRUITCOM, and that it is having some deleterious effects on performance and, hence, productivity. Sufficient medical evidence exists to point to the health concerns; the number of no-fault transfers attributable to stress also speaks forcefully to the issue. Individually perceived stress is telling in its overwhelming presence in the recruiter interviews and survey. Recruiters feel the stress, and people are being lost due to stress.

Finally, senior Navy managers have, through their own observations, reached the conclusion that the time has come to put stress under control. Adding to the urgency of that task are the views of the House Appropriations Committee. In its report on the defense appropriations bill, the Committee urged the services to relax goal-oriented recruiter evaluations somewhat, and in general, improve their quality of life, morale, and well-being. The Committee sees a link between the stress and pressures of the recruiter job and various forms of recruiter malpractice. With that, the Congress has acknowledged recruiting to be one of the "most stressful and difficult" jobs in the military ("House Panel Wants Less Stress," 21 August 1989).

With evidence in hand, it was time to get underway. COMNAVCRUITCOM chose to address stress through a coordinated attack that takes into account the mission, the human and other resources, and the working conditions of recruiting. Under development is a comprehensive stress management program to be implemented throughout NAVCRUITCOM to better assist recruiters to accomplish the Navy recruiting mission in the difficult years ahead.

References

Aunins, A. E., Sander, K. E., Giannetto, P. W., & Wilson, S. J. (in press). *Navy recruiter survey: Content analysis of free response data.* San Diego: Navy Personnel Research and Development Center.

Baker, H. G., Somer, E. P., & Murphy, D. J. (June 1989). *Navy recruiter survey: Management overview* (NPRDC Tech. Rep. 89-16). San Diego: Navy Personnel Research and Development Center.

House panel wants less stress in service recruitment. (21 August 1989). *Navy Times*, p. 8.

Recruiting Long-Range Plan Study Group (April 1989). Briefing on results of the Recruiter Survey Project. San Diego: Navy Personnel Research and Development Center.

Robertson, D. R. (1989). *Navy recruiter survey: Interview phase* (NPRDC Tech. Note 89-16). San Diego: Navy Personnel Research and Development Center.

Navy Recruiting: A Positive Approach to Improving Performance

Mr. Pat Lerro
Human Resources Research Organization (HumRRO)

Abstract

Based upon material from a top-down review of the Navy recruiting structure, this paper represents the perception of senior-level recruiting managers as to the kinds of research needed to help reduce stress upon Navy recruiters.

Military recruiting is a complex, highly visible process. Subject to a great deal of critical review from the Congress and the public, recruiting has generated a great deal of research. HumRRO identified over 1000 studies, conducted over the last 15 years, which deal with varying aspects of military recruiting. The application of these studies, and the research communities involved, can be seen throughout the recruiting process. Econometricians, operations researchers, and behavioral scientists have made significant contributions toward identifying, understanding, and influencing the youth market. IO psychologists and psychometricians have helped generate major changes in the areas of cognitive testing, enlistment standards, and job performance measurements. Specialists in manpower modeling have created state of the art automated systems to improve the recruiting aspects of the time-phased training pipeline flow, and to create complex models to perform trade-off analyses.

There have also been a number of significant studies done regarding the Navy recruiter. The record of application, however, regarding such research has not been very successful. It is ironic, that while all involved with recruiting agree that the recruiter is the linchpin to success in recruiting, the results of recruiter research have generally not been used to any great degree. The purpose of this paper is not to discuss the reasons for the low acceptance of recruiter research. Suffice to say that, in regard to Navy recruiters, there were a number of contributing factors. First, there was a perception that Navy recruiters were an unlimited resource; they could be replaced at any time. Second, unlike the other recruiting research, the application of many recruiter studies recommendations would have to have been applied by offices outside of the recruiting community; offices with differing priorities. Third, during the 1981 to 1986 period there did not appear to be any difficulties in achieving the recruiting mission. The perspective then, was why change something which seemed to be working.

The purpose of this paper is to explain the climate which placed tremendous pressure upon Navy recruiters, describe events which resulted in Navy leadership recognizing the need for change, and outline those elements of a research agenda which could facilitate needed change.

THE CLIMATE

There were three primary factors driving the need to reevaluate the Navy's recruiting force: An increasingly difficult recruiting environment;

the results of two independent studies and an IG report regarding Navy recruiting; and Congressional concerns.

The recruiting environment. Recruiting difficulties are no longer only the problem of military recruiters. A continually declining youth market, whose popluation shows growing deficiencies in comprehension and math, and in science and technical skill development, is creating nation-wide problems for public and private sector recruiters. The Army, and even the Air Force with its greatly reduced recruiting objectives, were feeling the impact of a more difficult market. Each of those Services compensated for these difficulties by adjusting resources, techniques, and product offerings. The Navy, however, remained fairly static in its support of the recruiting force over the same time period. This resulted in a less-than-competitive posture, which placed a great deal of pressure on Navy recruiters. Navy recruiters were forced to achieve mission on a day-to-day basis as opposed to being able to work with a six month lead time.

Comparing Navy recruiting quantitative results with those of the other Services indicated a last place performance in key areas of measurement. The percentage of high school diploma graduates and high aptitude youth entering the Navy relative to the other Services dropped significantly. At the other end of the spectrum, the Navy was accessing 75 percent of the Category IV recruits (the lowest acceptable test category permitted in the military). Further a review of the numbers showed that since FY 1986, the Navy actually had not recruited even the numbers of persons needed to meet new contract requirements. The decline in the Navy's recruiting posture finally lead to a call for an independent review of the Navy recruiting structure and program.

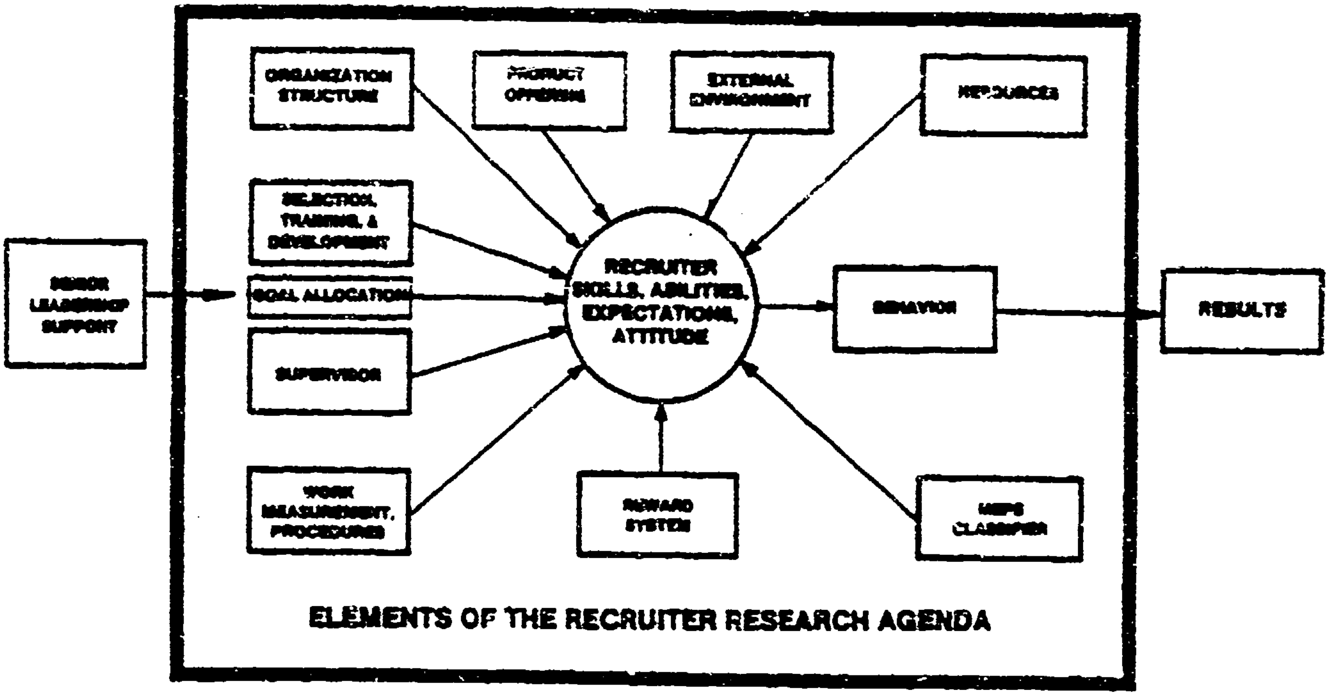
Independent reviews of Navy recruiting. There were three separate, simultaneous reviews of Navy recruiting: HumRRO was commissioned by the Office of the Deputy Chief of Navy Operations/ Program Planning; the Deputy Chief of Naval Operations/Manpower, Personnel, and Training commissioned an independent review headed by retired Vice Admiral Joseph Metcalf III; the Navy Inspector General inspected the Navy Recruiting Command. Although each group had differing objectives, their findings as to the causes of the problem were quite similar: Senior leadership had not actively supported the recruiting program; there was no master strategy nor focused management structure to guide the recruiting program; the Navy staff was not equipped to evaluate or defend recruiting resource requirements, and, the recruiting command (CNRC) was unable to develop programs needed to understand the wants and desires of the market, and was unable to develop the programs and incentives to attract that market.

CNRC met its minimal recruiting goals despite being at a distinct competitive disadvantage. However, that performance exacted a heavy human toll on CNRC's recruiters. Compounding the stress on the recruiter created by the poor competitive posture of the Navy, was the field leadership solution to not meeting recruiting quotas: simply replace the recruiter rather than examining the product, the training, or the local market. Some districts replaced half their recruiting force in a one year period of time. Such a questionable practice placed an even greater burden of stress on the remaining recruiters. It also gained Congressional attention.

Congressional concerns. The House Appropriations Committee, in its FY 1990 report, expressed concern that increasingly constrained market and economic climates may lead to undue pressures on recruiters to choose between maintaining their integrity at the cost of their careers or bending rules to make their goals. The Committee mandated a thorough review of recruiting policies and procedures to assure recruiting objectives can be achieved while affording recruiters a quality of life comparable to that experienced by most enlisted personnel. They also mandated a report to Congress, and promised hearings on the subject during the next session of Congress.

Identifying Needed Changes

If the Navy is to truly succeed in meeting its recruiting missions, without placing tremendous stress upon its recruiting force, there is a need to better understand the recruiter, and the impact of various elements which affect the recruiter. The following figure presents the various elements which affect recruiter performance, and briefly discusses those critical elements where further research could help reduce stress while helping the recruiter achieve mission.



ELEMENTS OF THE RECRUITER RESEARCH AGENDA

Recruiter selection. There are two types of research needed in this area: The first is to gain an increased understanding as to why sailors are extremely reluctant to volunteer for recruiting duty; and the second is to modify the current selection process.

A survey of sailor perceptions. Perhaps the greatest difficulty with recruiter selection is that relatively few sailors are interested in recruiting duty; it is clear that Navy recruiting is not perceived as a career-enhancing job. However, other than anecdotal comments, there are no real data which outline the depth and breadth of this problem. There is a need for a survey to identify the specific reasons for such reluctance, and a follow-on effort to overcome these objections.

Improve the recruiter selection process. The current recruiter selection process essentially just screens out individuals who are likely to fail, as opposed to a system comprised of more applicable standards that would select-in those individuals who have the greatest likelihood of success. Subsequently, many outstanding sailors selected for recruiting duty may not have the ability to be top-notch recruiters. Using the information obtained in earlier research as a base line, a new research effort should be implemented to develop a profile of the type of individual who has the greatest potential for success in recruiting. However, unlike earlier efforts, the research community should first find a sponsor who will ensure that the subsequent Battery be administered to a maximum number of job candidates.

Organizational structure. The Navy ratio of recruiters to zone supervisors is approximately 20 to 1. By comparison, the Air Force ratio is approximately 8 to 1, and the Army [if you include the company commander (equivalent to the zone supervisor), senior recruiter, and training NCO as supervisory personnel] is also about 8 to 1. One question for the researcher is, " With no additional manpower spaces, how can the structure be modified to improve the performance of a group of individuals, who have been provided only a modicum of training, and who currently receive minimum field training or supervisory assistance". From another perspective, research is needed to evaluate the role of the zone supervisor, and the pressures placed on that supervisor given the range of his/her responsibilities and his/her span of control.

Work measurement and procedures. The current measure of success in Navy recruiting is whether monthly mission was made. This places many recruiters in an almost untenable position where they are under constant pressure. It also sacrifices future success in order to achieve short term success. What is needed now is research from a bottom up perspective aimed at developing a criterion-based system to identify the key result areas and procedures which will help achieve those results. Such a research effort requires the direct involvement of experienced recruiters and recruiting managers, the application of industrial engineering techniques, and the involvement of behavioral scientists, operations researchers, and policy analysts.

Reward system. In this area, the key to effectiveness is an incentive system understood and perceived as attainable by the entire recruiting force. Navy recruiting officials have modified the reward system which was in effect last fiscal year. However, we believe that adjustment was only a first step. What is needed is a separate effort which allocates goals and provides rewards which take into account regional differences, including the level of difficulty.

Training and development. While recruiters represent the heart of the system, the training and development of the recruiter force represents the glue that binds the system into a cohesive whole. However, if training and development programs are imposed on recruiters and are interpreted as a disciplinary action or punishment for inability to perform, there will be little enthusiasm and little effective training. Additionally, if directive type training is imposed on recruiters who are overworked or under a great deal of stress, that training is likely to interfere with effective learning. The two factors above are a real problem to CNRC. As indicated in NPRDC's 1989 recruiter survey, over 65 percent of the recruiters perceived that the tab system, which should be part of a training base, is used as punishment; over a third of the force believed that fear and intimidation are the norm for supervisors and chief recruiters. A strong research effort is needed to determine the actual needs of the recruiting force, and then to determine how to effectively administer such training, given the very large span of control of Navy recruiting supervisors and the geographical dispersion of the recruiting force.

CONCLUSION

In any field, job performance represents the measurement of success, and the formula for job performance is:

Job Performance = Ability x Motivation x Opportunity

This formula is particularly critical in the recruiting environment. The combined reviews of three separate study groups, supported by the NPRDC survey of the Navy recruiting force, demonstrates marked deficiencies in the three factors which comprise this formula. These deficiencies have combined to place the Navy recruiting force under a great deal of pressure and subsequent stress. Substantial research is needed to identify effective long range solutions. Senior Navy leadership is providing the opportunity to develop a recruiter research agenda, and the motivation for such research is strong (to help solve operational difficulties as well as the need to satisfy congressional concerns).

Some may challenge the need to conduct such research due to the probability of greatly reduced recruiting objectives. We must understand that any reduction of recruiting objectives will also probably be met with reductions in recruiting resources. At the same time, some decision makers will try to raise quality standards given that the overall numbers are reduced. If the reduction in goal is met with a comparable reduction in recruiters or resources, while at the same time standards are tightened in a more competitive marketplace, the remaining recruiters will be forced to process even greater numbers of applicants per recruiter to achieve mission, thereby placing even further stress on the individual recruiter.

References

Lerro, P.A., & Barnes, J. (June 1989). Navy Recruiting: Planning for the Future (Report No. FR-PDR89-09). Alexandria, Va: Human Resources Research Organization.

Metcalf, J. (May 1989). Navy Recruiting - Metcalf Study. Unpublished Report for the Deputy Chief Naval Operations/Manpower, Personnel, and Training: Washington, D.C.

Major Findings Of The 1989 Navy Recruiter Survey

Laura E. Swirski
Emanuel P. Somer
Herbert George Baker, Ph.D.
Navy Personnel Research and Development Center

Abstract

This report presents an overview of results from a quality of work life survey administered to all production Navy recruiters. Ninety-four percent responded. Survey responses are discussed under the following topic areas: selection and training, goaling, career development, and quality of life.

Projections show a shrinking manpower pool from now through the first few years of this decade. Navy recruiting policy makers must decide whether to: (1) accept lower quality applicants, (2) pay the additional cost of competing for high quality youth, or (3) fail to recruit the number and quality of required recruits. In addition to the shrinking manpower pool, general unemployment remains low and civilian jobs are relatively plentiful in most areas of the nation. Competition for qualified people continues to increase. Educators and school counselors do not view military work as a desirable career option. This results in few youths considering the Navy or military as entry-level employment.

Despite these obstacles, the Navy recruiting force has achieved its goals in both the quantity and quality of accessions. Unfortunately, the results have been attained at great cost to the recruiters personally and often professionally. The quality of work life for recruiters has suffered greatly (Baker, Somer, & Murphy, 1989). The Commander, Navy Recruiting Command (CNRC). has initiated several short- as well as long-term efforts to address this situation. Major findings of the 1989 Navy Recruiter Survey relevant to four areas of organizational life which are impacted by recruiting duty are discussed below.

Results

Selection and Training

Recruiters perceive problems in connection with determining who is assigned to recruiting duty. The rigors of recruiting assignments demand that personnel selected to be recruters have the necessary characteristics that will sustain them, and that their family and financial situations do not hamper their performance. Relevant survey results indicate over 60 percent of recruiters questioned whether persons who have the proper personality and motivation are selected for recruiting (see Figure 1); 84 percent of the recruiters agreed that the Navy should develop a better way to select recruiter candidates. Nearly 75 of the recruiters were advised by former recruiters not to take an assignment to recruiting duty and 60 percent would decline to advise their friends to do so; 87 percent agreed that recruiting duty should be strictly voluntary (see Figure 2).

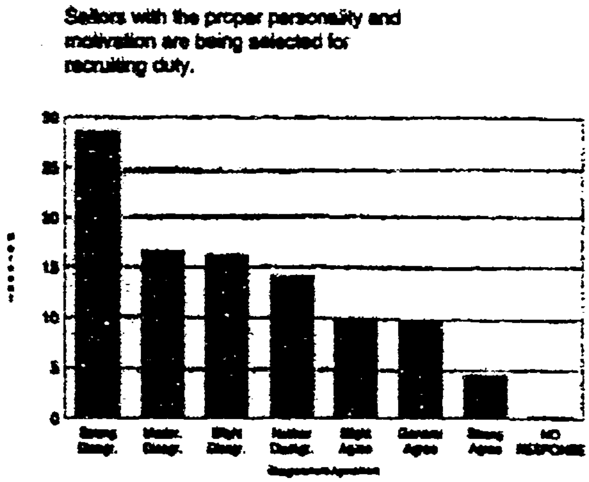


Figure 1.

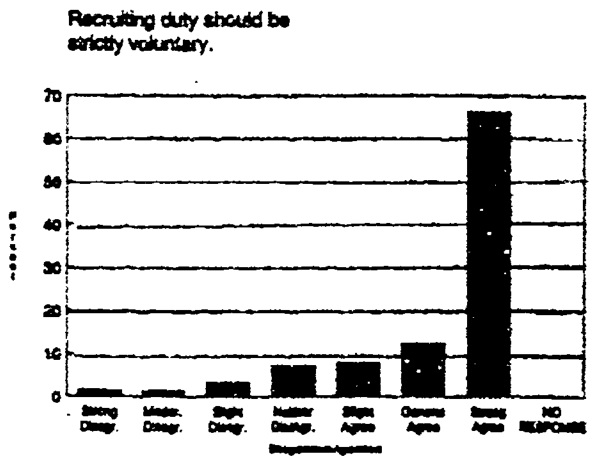


Figure 2.

There was a general impression that the training received is insufficient for those who are selected for recruiting duty. Respondents indicated that they were not prepared for the real world of recruiting or for the stresses and pressures of recruiting duty. They recommended realistic job previews to improve the quality of work life. There was also a strong consensus that sales training they had received was inadequate and that additional training would be helpful (see Figure 3). A majority responded that training which allows the experiencing of failure would be beneficial (see Figure 4).

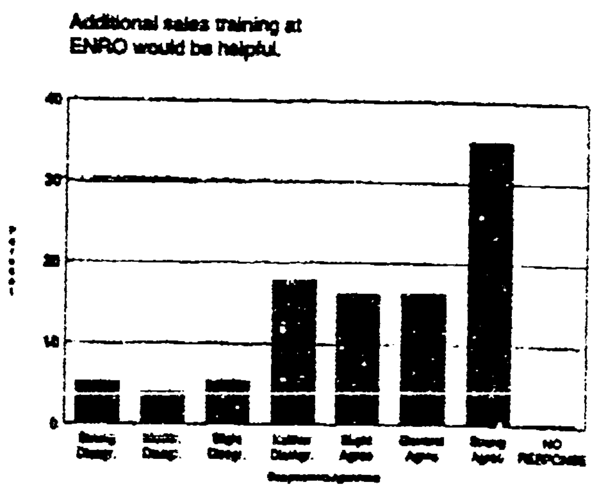


Figure 3.

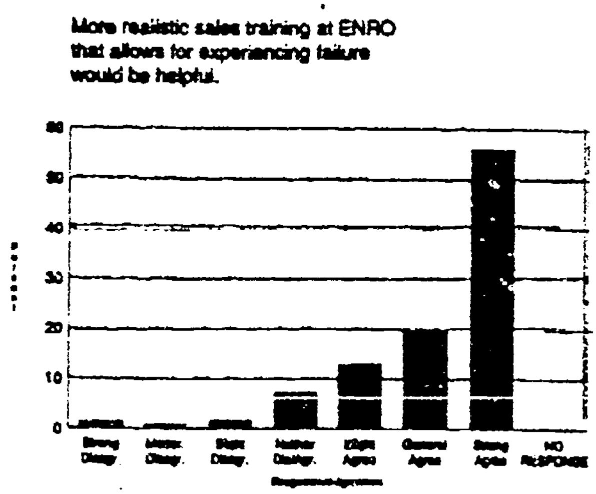


Figure 4.

Goaling

Recruiter success is measured primarily by whether or not they reach their goal. Unfortunately, nearly 75 percent of the respondents report that they do not have control over many of the conditions that determine if they reach their goal. Recruiters agreed that the 1-month goal period is too short (see Figure 5).

Eighty-five percent reported that they felt pressure to continue recruiting even after they had made their goal. No matter if their accomplishments one month, at the beginning of the next month the pressures returned. To ease the pressure on individual recruiters, recruiters strongly favored the team concept. Over 70 percent believed that station personnel should be organized into teams so that recruiters could specialize in the aspects of the job they perform best (see Figure 6).

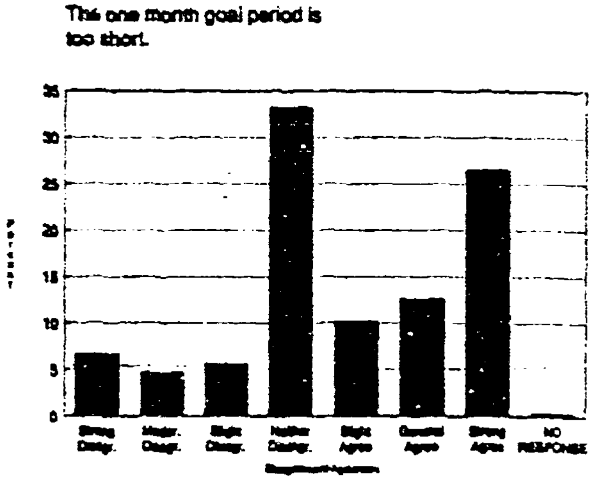


Figure 5.

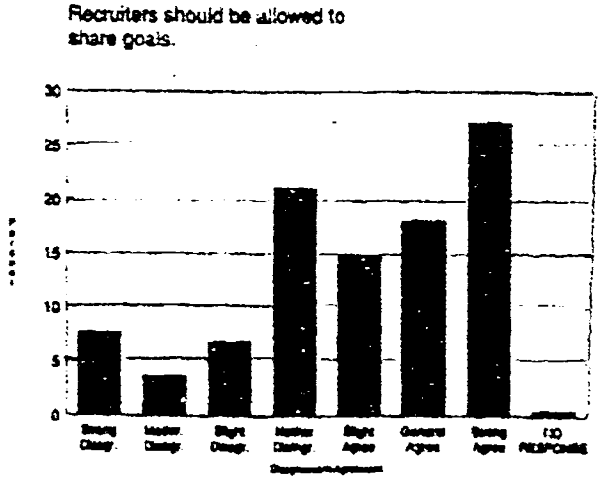


Figure 6.

Career Development

In the area of career development, recruiting falls short. Only 42 percent reported that recruiting duty would help them advance in the Navy. Many felt that recruiting would hurt their chances for advancement within their rating. One-third of the recruiters perceived their evaluations to be lower on recruiting duty than they were back in their occupational field. One-fourth felt their evaluations did not reflect the effort they put into their recruiting duties. About 70 percent agreed that, while on recruiting, job pressures have kept them from studying for their advancement exams, they did not have the time and material to study, and they should be relieved of goals to study for advancement. Recruiters ranked their first choice for improvements that receive extra credit at any selection or promotion board because of recruiting duty.

Recruiter recognition produced mixed responses. Sixty-four percent agreed that they received recognition for good job performance. However, over half claimed they received more recognition in the fleet than in recruiting duty. Nearly eighty-eight percent of respondents suggested that all of the time spent on recruiting duty should count toward sea duty.

A career ladder for recruiters is essentially nonexistent. A majority (72%) agreed that recruiters should have career development opportunities within recruiting after they have demonstrated their abilities.

Quality of Life

Job responsibilities, naving time to attend to personal matters, concerns about finances, and availability of support systems all contribute to the quality of life for Navy recruiters. About 80 percent of respondents reported much stress in their job and 79 percent agreed that recruiting was more stressful than other Navy jobs (see Figure 7).

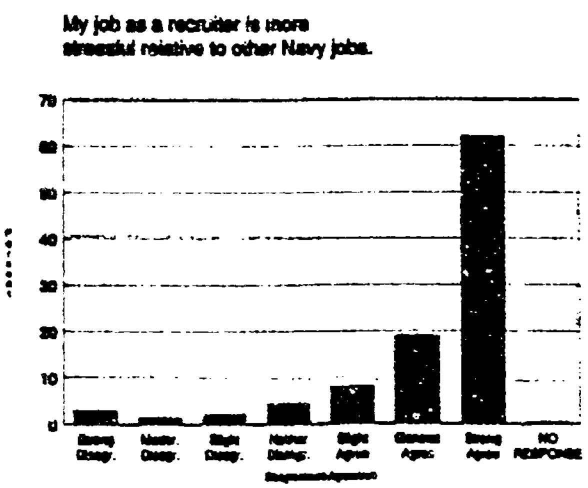


Figure 7.

Three-fourths disagreed that their work schedule left enough time to take care of personal business. Eight-five percent reported working 60 hours or more per week with nearly half of the respondents putting in 70 or more hours. Eighty-five percent disagreed with the statement that they were compensated for working holidays and extra hours.

Financial matters were also a source of concern. About 70 percent did not agree that the variable housing allowance was adequate to cover their living expenses. Personal expenses for medical care for self and family also appear to be a problem.

Sixty percent of the respondents disagreed with the statement that their recruiting district provides good quality counseling services related to personal and family problems. About 40 percent of the recruiters did not agree that their spouse, girlfriend, or boyfriend understood the pressures of being a recruiter.

Discussion

Despite the obstacles faced by the Navy recruiting force, goals for the quantity and quality of accessions have been achieved. However, the cost to recruiters has been great both personally and professionally. In an increasingly competitive employment market, the Navy must put forth the efforts necessary to select, place, and retain highly qualified individuals in the recruiting force. Attention to the issues that concern recruiters will impact both the productivity and quality of life of the Navy recruiting force.

References

Baker, H.G., Somer, E.P., & Murphy, D.J. (1989, June). Navy Recruiter Survey: Management Overview. (NPRDC Tech. Rep. 89-16). San Diego: Navy Personnel Research and Development Center.

Brief Cognitive Oriented Group Intervention for Treating Basic Trainee Adaptation Problems

Major Edwin L. Gerwell, Ph.D.
Edna R. Fiedler, Ph.D.
Wilford Hall USAF Medical Center

Abstract

This study reports on the effectiveness of brief cognitive behavioral oriented group therapy in attenuating minor adjustment problems in the Air Force basic training environment. Outcome data in terms of successful or nonsuccessful completion of training is reported.

Enlistee attrition during basic training has traditionally been considered a natural element of the training regimen. This supposition is based on the assumption that the young inductee comes to the training with inherent psychological and physical qualities which determine success or failure when faced with the calculated stressor of basic training. From time to time, this assumption has been called into question (Novaco, Cook, & Sarason, 1983; Buddin, 1988). Literature on stress and coping strongly sugggests that successful adaptation to various stressors, including actual combat, is as much a function of how individuals are prepared for the stressor as their inherent adaptative qualities. Research evidence suggests that simple and effective psychological procedures exist which could significantly reduce attrition in otherwise psychologically healthy individuals (Butcher, & Maudel, 1976). As defense budgets shrink there is increasing support for gaining the most from the personnel dollar and reducing or eliminating unnecessary attrition of qualified enlistees. The current study supports the concept that short and efficient psychological methods exist which can help support this trend toward personnel cost conservation.

Method

Subjects

Subjects were 436 basic trainees who were previously seen for command directed evaluations at the Behavioral Analysis Service, a dedicated mental health facility. Subjects were evaluated over a five month period and determined to be capable of returning to duty. The evaluations indicated that these individuals would benefit from brief psychological intervention to increase their coping skills upon return to training.

Apparatus

Group sessions were provided at the Behavioral Analysis Facility located in close proximity to the training area at Lackland Air Force Base. This

---

The authors express their appreciation to Tom Swanson for his help in statistical analysis and Elizabeth A. Rupprecht for her editoral input. The ideas presented in this paper do not necessarily reflect the official policy of Wilford Hall or the Department of Defense.

provided a comfortable relaxed and private environment. Each group of approximately 15 individuals was led by a supervising Ph.D. level psychologist and pre-doctoral resident. A psychiatry technician was also present to handle administrative issues and maintain required medical notes.

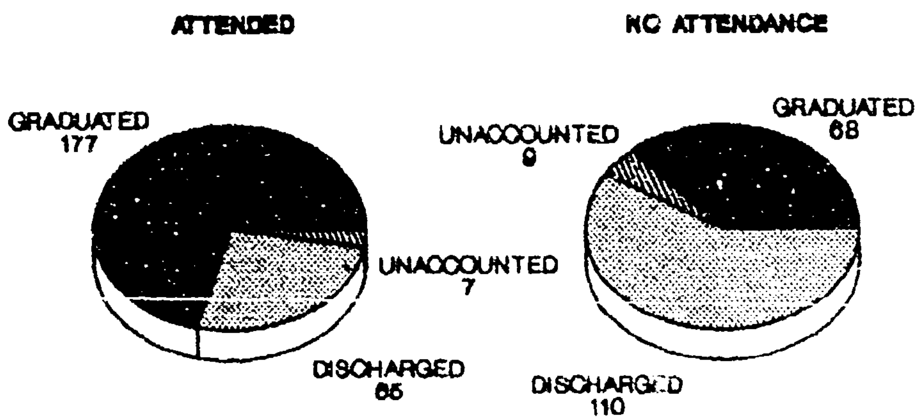
Procedure

Basic trainees were first evaluated at the Behavioral Analysis Service, based on a commander directed referral, and determined to have a nondisqualifying diagnosis but with some degree of coping problems. Each was provided an appointment and recommendation for group attendance. Attendance was determined by acquiescence of squadron commander. Subjects were routinely checked for whether they had attended group and their graduation or discharge status. Group sessions consisted of two one hour sessions each week. Each session involved brief inputs from each member, limited problem solving and supportive inputs from other members and a combination of psychoeducational and cognitive coping inputs from group leaders (Meichenbaum, 1977).

Results

Analysis of outcome data for the subject groups collapsed over five months contained in figure 1 clearly supports the conclusion that group attendance contributed significantly to success in basic military training. In the group of individuals who attended the group, 177 (71.1%) graduated versus 68 (36.4%) in the group who did not attend. The 16 total subjects unaccounted for, in this study, are considered insufficient to influence the global effect for attendance. No accounting was made of the nature of discharges as these are generally of an administrative type. It is also possible a very small number of medical and other discharges are contained in the discharge numbers of both groups. A chi square analysis showed that attendance was significantly related to graduation and no attendance was significantly related to discharge ($p. < .005$).

Figure 1
Group Attendance and Training Outcome



Discussion

One conclusion drawn from the current study is that a relatively small investment of intervention resources can produce a major impact on individuals who are experiencing a significant crisis situation. Trainees attending the group averaged two to four sessions. Total active involvement for each individual is estimated to be no more than five minutes per session. This latter observation would suggest that modeling, mutual support and other factors known to exist in group situations may contribute to a large degree toward the outcome found in this study. This suggests future research to determine specific intervention elements which differentially affect this type of crisis intervention. Although group members were determined to be qualified to return to duty and training the question arises as to their capacity to complete a full enlistment (McGraw & Bearden, 1988). Although this question cannot be answered in this study, it is being pursued and will be addressed in a follow-up study on this population.

It is well known that the acute crisis element involved in basic training extends into multiple other situations which are common in the military including war, dislocation, natural disasters, medical evacuation teams, survival school instructor and military instructor duty and the like. The impact of stressors arising from such sources also extend to the families of active duty personnel. The negative impact of these situations are attrition, increased medical costs, diminished morale and readiness. This study suggests that modest intervention efforts could make a major impact on the aforementioned factors. The fact that calculated savings in attrition costs resulting from the group intervention at the Behavioral Analysis Service are close to two million dollars, lends strong support to such a conclusion. This alone should be a strong incentive for further research into the application of stress interventions for military operations.

References

Buddin, R. (1988). Trends in Attrition of High-Quality Military Recruits. Santa Monica, CA: RAND Corporation.

Butcher, J.N. & Maudel G.R. (1976). Crisis intervention in I. Weiner (ED.) Clinical Methods in Psychology. New York: Wiley Press.

McGraw, R.K. & Bearden, D.L. (1988). Motivational and Demographic Factors in Failure to Adapt. Military Medicine, 153, 325-328.

Meichenbaum, D. (1977). Cognitive-behavioral modification. New York: Plenum Press.

Novaco, R.W., Cook, T.M. & Sarason, I.G. (1983). Military Recruit Training: An Arena for Stress-Coping Skill. In D. Meichenbaum & M. Jarenko (Eds.). Stress Reduction and Prevention, (pp. 377-418). New York: Plenum Press.

Interactive Videodisc:
Using technology to improve human relations skills

David B. Porter, D. Phil
Lisa M. Ambre
United States Air Force Academy

Abstract

Interactive Videodisc (IVD) technology was tested as a valid teaching technique in conjunction with an already existing interpersonal relationship skills training program. A group of Air Force cadets used the IVD system in a quasi-clinical setting, and their learning was measured in terms of skill, motivation, and knowledge. The results show that the IVD system exhibited a positive influence in all three areas.

Word processing and spread sheets have already demonstrated the impact of computers on the industrial community. Now there is an explosion of growth in computerized education and the leading edge of this explosion combines both computer and video disc technology. When these technologies are yoked together by Interactive Videodisc software, they form "the ultimate educational tool." These systems are revolutionizing training and education in today's business world (Currier, 1983). Their growth is likely to increase as computers and video systems become cheaper, more powerful and more widely available. It is important that this new technology be better understood.

Bainbridge (1988) has cataloged an impressive set of IVD successes. In a University of West Florida program, students using the videodisc showed a 25 percent reduction in training time and a 16 percent increase in the number of students who passed a standard written test compared to those exposed to traditional teaching methods. Similar results were found by Davis in the Annenberg project. She found that students took less time to complete courses and achieved the same or higher scores using interactive videodisc. In addition, a National Science Foundation study found college students learning biology on an interactive videodisc system took 37 percent less time to complete the course than those in the traditional classes and showed significantly higher scores on the post-test.

The use of interactive videodisc instruction has not been limited to academic settings. Computer based training (CBT) has become increasingly popular in private industry and goverment agencies as well. In 1988, the Air Force Commissioning and Education Committee (CEC) endorsed a proposal to examine the feasibility of using IVD technology to enhance professional officer development (Eagle Corp., 1988). The committee wanted to use IVD technology as a common tool for teaching leadership skills to officer candidates from three different commissioning sources (the Air Force Academy, Officer Training School, and Reserve Officer Training School). The committee decided to investigate the educational effects of this technology before investing development funds. Two key aspects of this investigation where 1) to find the best available off-the-shelf software, and 2) to explore the possibility of integrating group process with this new technology.

The purpose of our study is to further the CEC research on the use of IVD technology as an effective educational tool, and to look at the possible advantages of using student teams to process and enhance the IVD program. An IVD program developed for teaching interpersonal relation skills developed by Wilson Learning had been selected and extensively tested (Porter and Guerrero, 1989; Porter, Burtley, & Guerrero, 1990). This IVD program was integrated into the Skills Training Program at the United States Air Force Academy. The purpose of this eight-week program was for cadets identified by their commanders as having interpersonal skills deficiencies to increase their interpersonal effectiveness. Cadets in the program meet weekly for about two hours to work with a counselor to develop their interpersonal skills. As with previous studies (Porter, Burtley, & Guerrero, 1989) the Nominal Group Technique was employed as a way of using group interaction to enhance learning outcomes. This study involved individual self-assessment, group and individual performance, and expressed satisfaction. Data from this quasi-clinical group was compared to data from a variety of other groups that had used the same IVD software and nominal group technique. One particularly important aspect of the IVD program is that it provides immediate feedback. Yalom (1985) claims that through feedback and self-observation, one is able to understand his/her own behaviors better and is able to appreciate the impact of that behavior on others.

Method

Subjects

Three female and three male cadets ranging in age from 18 to 22, Academy freshmen through seniors, served as subjects. The members of this group had volunteered for the Skills Training Program. The control group data was used for comparison came from the previous studies with cadet and officer trainees (Porter & Guerrero, 1989). Comparison group subjects had used the same techniques and technology that were employed within this study.

Apparatus

A portable Sony View Interactive video system with software from was Wilson Learning's Interpersonal Relations: Interacting with Others was used. Three study-practice units in this package covering developing rapport, listening to others, and displaying sensitivity were presented during the first session. Each unit required approximately 15 minutes to complete and included instruction, demonstration, evaluation, and feedback. A 35 minute comprehensive grand practice served as the criterion task. The grand practice assessed the skills of the students on all three of the study-practice units. In addition, a workbook containing 15 self-assessment questions, instructions on the nominal group technique, spaces to record individual and group responses, and a program critique was given to each subject.

Procedure

Two groups (one of two subjects and the other of four subjects) were used to conduct the training. This was accomplished during two sessions separated by approximately one week. The two groups did not interact. On the first day, subjects were given an introduction by the first author then completed the interpersonal relation skills self-assessment questions and were given brief equipment operating instructions. Next, the students completed the three study-practice units using the nominal group technique: each time the group arrived at a decision point each individual marked an answer in his/her workbook. Only after everyone had marked their individual response, could the group discuss the problem and reach consensus on the appropriate response. This choice was marked in their workbooks and input to the computer. During the second session, the subjects again followed the same procedure. Subjects then individually filled out the IVD program evaluation. The questions on this evaluation were designed to measure three different outcomes: skill, knowledge, and enjoyment. Questions tested satisfaction on each individual category as well as overlapping areas. Responses ranged from 1 representing strong disagreement to 5 representing strong agreement.

The group containing the four cadets was also timed by the second author to more precisely examine the temporal characteristics of each cadet's participation in the group discussion. The sequence of discussion was recorded as well.

Results

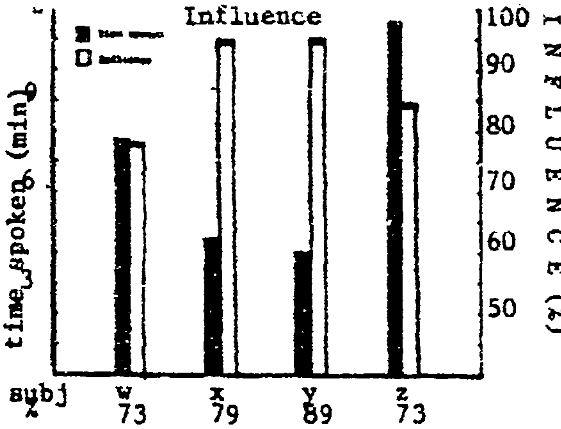
The demographic data revealed that the groups had both low grade point averages (X=2.25, s=.139) as well as low military performance averages (X=2.11, s=.139). Cadet averages on both of these measures are about 2.75 with a standard deviation of .50. These subjects' low scores document low performance, however they do not necessarily validate their commander's attributions that the cadets' deficiencies were due to a lack of interpersonal skills.

Three types of measurements were obtained from this study. They all dealt with interpersonal skills; specifically self-assessment, performance, and program satisfaction. The means and standard deviations obtained for each of these three catagories were similar to those found by Porter and Guerrero (1989). On a scale of 1 to 5 with 1 being below average and 5 being outstanding, the scores were as follows: building rapport (X=3.53, s=0.97), listening to others (X=3.35, s=0.88), displaying sensitivity (X=3.73, s=1.11). These results indicate that subjects saw themselves as having interpersonal skills that were between above average (3) and well above average (4). Although these means did not differ significantly from the control group, all were slightly more positive. Individuals enrolled in the program apparently did not share their respective commander's assesment of their interpersonal deficiencies.

The grand practice provided the performance criterion. The number of correct responses made by each individual before and after group interaction was compared. An average of 15.5 (s=1.4) of subjects' 19 initial responses were correct. After group discussion and consensus, the average number correct was 17.0 out of 19. When these scores are compared to those obtained from other groups (Porter & Guerrero, 1989) they are slightly superior. The subjects' perceived interpersonal deficiencies might not be perceptual, but could reflect inactment (behavior) difficulty. The group's 89.5% accuracy reflects a substantial improvement over their average individual accuracy of 81.5%. The group's influence had a positive effect on performance. On 10 of the 19 questions, at least one subject changed his/her individual answer to match the appropriate choice.

The group of 4 cadets was also measured individually in terms of the amount of time spoken, the percentage of times the individual choice matched the eventual group choice, and the percentage of times the group choice matched their individual choice given that they spoke last.

Fig 1. Subject Speaking Time and Influence



It is particularly interesting that every time subject X or Y spoke last the group decision matched their respective individual choices. This was not true for subject W or Z (Fig. 1). Ironically, the two cadets who spoke least (less than half as much as their colleagues) exerted the greatest influence on the group decision. Subjects' affective reaction to the training was measured by their ratings on the IVD evaluation at the end of the experiment. High scores show that the students were very satisfied with the program. No significant differences were noted between the three different outcomes (enjoyment, skill, and knowledge), but this might reflect a ceiling affect due to the extremely positive ratings.

Discussion

These results are generally consistent with those from previous studies (Porter & Guerrero, 1989; Porter, Burtley, & Guerrero, 1989) and provide some support for using this technology with a group which had been identified as lacking interpersonal skills. As predicted, the Interactive Videodisc system was an effective training tool. Student's were extremely satisfied and group interaction increased performance and may also have provided useful support. From the results, we know that in the group of 4 cadets, whenever there was dissension between the individuals' choices, it was always 3 to 1. This implies that when one of the three spoke to the group they were not only trying to influence the loner but were positively reinforcing the other two as well. This reinforcement may account for the high satisfaction scores obtained on the evaluation. As Yalom (1985) suggests, the immediate object of feedback that the program offered may have also been a critical contributing factor.

There were several other contributions the IVD sessions made to the Skills Training Program which were less obvious. The high technology equipment and excellent commercial program clearly communicated institutional committment to the subjects and the importance it placed on developing interpersonal skills. It also provided a useful variation on the usual group counseling format. Because the IVD program provided both structure and direction, counselors could gradually withdraw and more objectively observe subjects' interpersonal skills. Perhaps most importantly, subjects were able to have some fun while they learned something about interpersonal skills.

Author Notes

Views expressed in this paper are those of the authors and do not necessarily represent the view of the U.S. Air Force or any of its agencies. The authors gratefully acknowledge the support of the Frank J. Seiler Research Laboratory.

References

Bainbridge, S. (1988). Efficacy of the interactive management skills training program design of research and results to date, presented to the Society for Applied Learning Technology, Annual Conference, Orlando, Florida.

Currier, R. L. (1983, November). Interactive videodisc learning systems, <u>High Technology</u>, pp. 51-59.

Eagle Technology, Inc. (1988). <u>Computer Network to Support OTS Information Systems and Training - USAF Officer Training School</u>. (San Antonio: Eagle Technology Inc. AF contract No. F41689-86-D0009).

Hassett, J. & Dukes, S. (1986, September) The new employee trainer: A floppy disc, <u>Psychology Today</u>, pp. 30-36.

Porter, D. & Guerrero, J. C. (1989, March). Interactive video disc: Process description and initial evaluation, presented to the Southwest Division Academy of Management, Thirty-first Annual Meeting, New Orleans, Louisiana.

Porter, D., Burtley, B., & Guerrero, J.C. (1989). Human relations skills and interactive video disc: Using student groups to enhance educational outcomes.

Yalom, I. (1985). <u>The theory and practice of group psychotherapy</u>. New York: Basic Books, Inc.

Who Follows the Rules when Nobody is Watching: Predicting Ethical Behavior among Academy Graduates

Earl H. Potter III, Ph.D.
Rex J. Blake, M.A.

U.S. Coast Guard Academy

Abstract

The service academies of the United States share the common goal of graduating commissioned officers who adhere to standards of conduct that are, in turn, based on broad concepts of honor and duty. In practice, meeting this goal should mean that graduates adhere to rules and regulations in both public situations, where their behavior is subject to scrutiny and evaluation, and private situations, where "no one is watching." This study examines the "rule keeping" behavior of graduates of the U. S. Coast Guard Academy. The California Psychological Inventory is used to predict the degree to which graduates follow the regulations in private. Rule keeping behavior is related to assessed performance of these officers during their time as cadets and during the first two years following their graduation from the Academy. Finally, the implications of these findings for the development of our future leaders is considered.

Each of the service academies is characterized by an exhaustive system of regulations. These range from rules anticipating those the graduates will encounter later in the Uniform Code of Military Justice (absence without leave, disobedience of orders, disrespect to seniors, etc.) to extraordinarily explicit and detailed rules addressing the minutae of day to day life (e.g., regulations stipulating the precise manner of folding underwear and socks and their correct placement in the drawers of the cadets' chiffoniers). The service academies each have promulgated an additional set of rules that emphasize adherence to a higher standard of ethical behavior: their "honor codes."

If detected, violation of any of these rules results in punishment. Although the various academies differ somewhat in their administration of discipline, sanctions for rule breakers typically include "demerits," extra work assignments, restriction to the academy grounds, and dismissal from the academy. At the Coast Guard Academy, the total number of demerits received by each cadet is tabulated and becomes a significant part of his/her overall military evaluation; accumulation of a sufficient number of demerits (even if these are due to a series of minor infractions) can eventually result in a cadet's dismissal from the academy.

From the emphasis placed on these systems of regulations, it may be inferred that rule adhering behavior is highly valued. The service academies of the United States share the common goal of graduating commissioned officers who adhere to well defined standards of conduct. In practice, meeting this goal should mean that graduates adhere to rules and regulations in public situations, where their behavior is subject to scrutiny and evaluation, and private situations, where "no one is watching." We draw the reader's attention, however, to the words that opened the preceding paragraph: if detected. It seems at least two possible outcomes are likely to result from this relatively inflexible and demanding system of rules; at the end of four years, the academies graduate either commissioned officers who adhere to the rules or officers who break the rules and are able to avoid detection.

An index based solely on the number of recorded infractions would fail to distinguish between these groups. This paper represents an effort to ascertain the actual degree to which academy graduates adhere to rules. Secondly, it examines the strength of the relationship between rule adherence at the academy and later performance of these junior officers. Finally, this study investigates the degree to which individual differences existing at entry are predictive of rule adhering behavior during the ensuing four years.

Method

Subjects

The Class of 1986 numbered 250 when they entered the academy in July of 1982. Four years later 138 graduated (128 of which were U. S. nationals). Just prior to graduation we conducted a structured interview of 76 cadets selected at random from the class. In the summer of 1988 we asked each of the 128 U.S. graduates to participate in a follow-up study; 120 of these agreed to participate and to date we have obtained complete performance records for 100. The majority of the incomplete records are due to lags in the performance appraisal system which seem to have no systematic pattern that might bias the results of the study. Thirteen of the graduates were women. Ninety-four percent of the graduates were 21 or 22. Only five were minorities. A complete data set for this study was obtained for 42 subjects.

Measures

Undergraduate Performance. The measure of undergraduate performance was class rank which is based on a composite of academic (75%) and military (25%) performance. While not pure, this criterion is much like the job performance criteria used in job selection studies. It is a mix of objective performance and style.

Job Performance. The Coast Guard assesses the performance of junior officers every six months using a 23 item behaviorally-anchored rating scale divided into six major areas: performance of duties, interpersonal relations, leadership skills, communication skills, personal qualities, and effectiveness in representing the Coast Guard. Each major area is followed by a narrative and the whole report is summarized in a comment of the officer's leadership and potential for positions of higher responsibility.

We reviewed the performance reports which each officer received during the two year initial assignment. Four distinct patterns emerged. First, an officer might start out strong and continue to perform at a high level. This officer was either seen as a natural leader with great value to the service or seen as unique in the rater's experience and very special. Second, an officer might start slowly making mistakes and demonstrating some weaknesses but develop rapidly during the course of the two years finishing as a strong contributor to the effectiveness of the unit. Third, an officer might be a steady, dependable but unremarkable performer who was an asset but not the key to success. Fourth, an officer might demonstrate weakness that persisted over the two years. For some the evaluator's judgement might be that the officer needed further development or would be successful in a less demanding assignment. These were marginal performers. Others were judged to be too weak for promotion. These officers failed to be promoted from Ensign to Lieutenant (junior grade) at the end of the usual 18 month trial period. Job performance for the two years immediately following graduation was, therefore, summarized by the following scale:

5. (N=8) Comments include "the best I've ever seen," "a rising star," "innovations will benefit the entire service." Demonstrated strength from the start, no weaknesses noted.

4. (N=46) Strong from the start, excellent initiative and stamina, "maturity unusual for an junior officer," highly recommended for command, but not "the best ever."

3. (N=29) "Solid performer," may have minor difficulties at the start but excellent record overall by second year, may be recommended for command but without enthusiasm, reports are positive but lack superlatives and excitement.

2. (N=16) Persistent flaws that impair performance, significant command attention needed to get the officer's attention, comments include "slowly improving," "recommendation for command withheld pending further growth," "recommended for responsible positions ashore." Not recommended for command.

1. (N=7) Failed to be selected for promotion to Lieutenant (jg) or evaluator expresses serious doubts that this officer will "make it," comments clearly convey negative evaluation, e.g., "this officer is lazy."

California Psychological Inventory. Each student entering the Coast Guard Academy is administered an extensive battery of personality inventories within six weeks of arrival. The results of the inventories are not used for screening or placement, but are retained for purposes of individual counseling and institutional research. For several years, the standard battery has included the California Psychological Inventory (Gough, 1957, 1966, 1987a; Megargee, 1972). The class of 1986 was administered the 480-item version of the CPI in the summer of 1982. Scores were obtained for each student on the eighteen standard "folk concept" scales and on a set of newly developed "structural" scales or "vectors"--three orthogonal scales designed to measure the major interpersonal themes underlying the folk concepts. The structural scales were scored on the original 480-item form according to instructions supplied by Gough in a personal communication.

Gough describes the first structural scale, v.1, as an index of interpersonal orientation. Low scorers on v.1 are described as "active, alert, involved, and extraverted" while high scorers are "more subdued, private, detached, and introverted". (Gough, 1987b) Among the folk concept scales, it correlates highly with Dominance, Capacity for Status, Sociability, Social Presence, Self Acceptance (correlations range from -.78 to -.50).

According to Gough, v.2 is a measure of the respondent's normative orientation. "This continuum goes from a norm-doubting, skeptical, and non-conformist perspective at one pole, to a norm-accepting, acquiescent, and rule-compliant orientation at the other." (Gough, 1987b). Its highest positive correlations among the folk concept scales are with Socialization, Responsibility, Achievement via Conformance, and Good Impression (correlations range from .67 to .52). It also correlates relatively highly, though negatively, with Flexibility (r=-.54 for males and r=-.50 for females).

Gough offers v.3 as a measure of "a range of fulfillment or realization ... proceeding from self-defeating and inadequate resolution of problems and potentialities specific to that type, up to an actualized and ego-enhancing integration." (Gough, 1988, p. 2) It correlates highly with virtually every one of the folk concept scales except Femininity (r=.02 for men, r=-.13 for women) and Communality (r=.27 for men, r=.30 for women). For the other scales, the correlations range from r=.87 (for Tolerance and Achievement via Independence) to r=.40 (for Flexibility).

Rule Keeping Behavior. Coast Guard Academy regulations are divided into three categories. The least "offensive" transgressions, Class III, including such things as a poor shoeshine and clothes improperly folded. Class II offenses include missing class, inattention to duty and gross deficiencies in appearance. In this study we asked cadets in a structured interview to identify which of the 41 listed Class I offenses they had committed without "getting caught". Class I offenses are the most serious offenses and include, for example, cheating, falsifying records, AWOL and sexual misconduct.

Results

The number of separate Class I offenses identified ranged from 1 to 25 with a mean of 7.5. There are only two significant intercorrelations among all of the measures. The first is between the CPI scale v.2 and the number of self reported Class I offenses ($r = -.60$). The second is the correlation between Class Rank and Performance (oer) after graduation ($r = -.29$). None of the other relationships--neither the relationship between the number of Class I offenses and overall academy performance nor the relationships between the number of Class I offenses and performance after graduation--were significant.

Table 1

	v.1	v.2	v.3	oer	Rank
Class I	-.15	-.60**	-.17	.09	.10
Class Rank	-.05	-.05	-.08	-.29*	

* $p < .01$
** $p < .001$

Discussion

The magnitude of the correlation between v.2 and the number of reported Class I offenses is not surprising. The normative orientation scale was developed to measure a factor that has emerged on virtually all factor analyses of the CPI. Its constituent scales generally reflect such themes as social conformity and self discipline. The number of Class I offenses provides a good measure of the degree to which the students abide by the rules of the organization or, conversely, reject those rules. More surprising, perhaps, is the lack of a relationship between the Class I offenses and overall performance, assessed either during the cadets' time at the Academy or after their graduation.

The fact that the number of Class I offenses committed is predictable but not predictive raises a host of questions concerning the match between our standards and our objectives. Perhaps the most troubling question concerns the meaning of these findings. Are we raising a generation of officers who conform in public but whose personal standards are at odds with the norms and values of the service--and what if we are?

References

Gough, H. G. (1957). Manual for the California Psychological Inventory. Palo Alto, CA: Consulting Psychologists Press, Inc.

Gough, H. G. (1968). An interpreter's syllabus for the California Psychological Inventory. In P. McReynolds (Ed.), Advances in Psychological Assessment (Vol 1). Palo Alto, California: Science and Behavior Books.

Gough, H. G. (1987a). California Psychological Inventory Administrator's Guide. Palo Alto, California: Consulting Psychologists Press.

Gough, H. G. (1987b). Some implications for managerial style of interpersonal and normative orientations. Paper presented at the symposium, New perspectives on personality and job performance. Ronald C. Page, Chair, American Psychological Association meeting, New York City, NY.

Gough, H. G. (1988). Testing for leadership with the California Psychological Inventory. Paper presented at the Center for Creative Leadership Conference on Psychological Measures and Leadership, San Antonio, Texas.

Megargee, E. I. (1972). The California Psychological Inventory Handbook. San Francisco, California: Jossey-Bass.

An Examination of Job Satisfaction and Career Intentions Among Air Force Officers, Enlisted Personnel, and Civilian Employees

Robert E. Niebuhr
Auburn University

Lt. Col. Mickey R. Dansby
Defense Equal Opportunity Management Institute

Abstract

A study of job satisfaction and career intentions of more than 30,000 Air Force personnel provided support for traditional models of turnover as well as indicating several significant differences between military and civilian subgroups.

High attrition rates in the Air Force (Finstuen & Berry, 1981; Ward & Tan, 1985) have placed increased emphasis on examining those factors which may provide additional insight into understanding the underlying attitudes of individuals who either remain in or leave the military. Of particular concern are the factors in their environment which may eventually cause them to be attracted away from their present situation or repelled away due to dissatisfaction with some aspect of their job (Niebuhr & Dansby, 1986).

In examining turnover models, research has focused on the impact of job satisfaction and career intentions relative to turnover decisions (Steers & Mowday, 1981). Recent path analysis research suggests that personal and organizational characteristics appear to influence job satisfaction, subsequently affecting organizational commitment, career intentions, and eventually actual turnover (Michaels & Spector, 1982; Williams & Hazer, 1986). Both job satisfaction and organizational commitment appear to be intervening variables in the turnover process, and it is apparent from the above studies and others (Martin, 1980; Steel & Ovalle, 1984) that the best predictor of turnover is the employee's intention to leave. Within the military setting, career intention is also the strongest predictor of reenlistment decisions (Ham & Hulin, 1981).

Two of the antecedent attitudinal variables examined in these recent path-analytic studies are perceived job characteristics and leadership. Reviews of research have generally supported the models that suggest that both job characteristics (Loher, Noe, Moeller, & Fitzgerald, 1985) and leadership (Bass, 1981) affect both job attitudes and behavior. Tasks which are perceived to be more motivating (due to the presence of skill variety, antonomy, feedback, task identity, and task significance) and situations in which the leader is more supportive suggest higher levels of satisfaction. Turnover models recognize that, over time, job satisfaction may be related (directly or indirectly) to turnover intentions and turnover actions.

Organization tenure (length of time in the organization) has been shown previously to be positively related to job satisfaction (Herman, Dunan, & Hulin, 1975; Spenser & Steers, 1981) and is viewed as inversely related to the propensity to search for new job opportunities (Steers & Mowday, 1981).

Based on support for current models of turnover it would be anticipated that motivating job characteristics, supportive leadership, and seniority would be positively correlated with job satisfaction, and to a lesser degree, with intentions to make the military a career. In a review of turnover studies Steel and Ovalle (1984) found that studies on military samples consistently resulted in greater attitude-turnover relationships than studies in civilian populations. Consequently, it would be predicted that the correlates with job satisfaction and career intentions for both officers and enlisted personnel would exhibit stronger relations than those of the civilian employees.

Method

Data for the present study came from the Leadership and Management Development Center (LMDC) Organizational Assessment Package (OAP) data base of almost 300,000 survey responses collected in LMDC management consultation visits to Air Force organizations. The surveys were administered from October 81 to July 85 as a census of organizations (usually wing size or equivalent) at 65 bases or sites where LMDC was invited to consult. Surveys were administered in group sessions and respondents were assured of individual anonymity of responses. Only LMDC consultants handled completed survey response sheets. Respondents for the present study were a 31,980-person segment of the data base, chosen to provide a representative sample of functional areas (operations, maintenance, and resources) and personnel categories (officer, enlisted, and civilian).

The OAP provides an Overall Leadership scale of nineteen items. The Task Motivation scale is comprised of five subscales and calculated in a multiplicative model developed by Hackman and Oldham (1976). The Job Satisfaction OAP factor consists of seven items and has been shown to be reliable (Short, 1985). Career Intention was measured using a single OAP item, and Seniority was derived from the demographic section of the OAP.

Results

Descriptive statistics, scale reliabilities (where appropriate), and Pearson product-moment correlations among the variables are presented below:

Table 1

Descriptive Statistics, Reliabilities, and Intercorrelations Among the Study Variables

	M	SD	1	2	3	4
1. Overall Leadership	4.79	1.46	(.97)			
2. Job Characteristics	109.56	65.79	.33	(.75)		
3. Job Satisfaction	5.07	1.16	.50	.50	(.81)	
4. Career Intention	3.74	1.31	.15	.27	.30	
5. Seniority	4.88	1.87	.01	.25	.10	.48

Reliability coefficients are in parenthesis in the diagonal. All correlation coefficients above .02 are significant at the .001 level.

Table 2 provides the results of the correlational analysis between job satisfaction and career intention for the three personnel categories (officers, enlisted, and civilian), while Table 3 details the analysis of the correlates under study for both of these variables.

Table 2

Correlations Between Job Satisfaction and Career Intention

Officers	.28
Enlisted Personnel	.29
Civilians	.24

Table 3

Correlates of Job Satisfaction and Career Intention

	Correlates with Job Satisfaction		
	Leadership	Job Characteristics	Seniority
Officers	.49	.53	.05
Enlisted	.49	.49	.08
Civilians	.54	.40	.08

	Correlates with Career Intentions		
	Leadership	Job Characteristics	Seniority
Officers	.14	.23	.43
Enlisted	.15	.25	.48
Civilians	.10	.17	.17

Discussion

The results from analysis of the data did provide partial support for the first hypothesis by finding that leadership and job characteristics, two of the three correlates examined, were more positively correlated with job satisfaction than with career intention. Seniority, however, was more related to career intention than to job satisfaction, possibly a result of the strong pension benefits available to military retirees and the recognition that an "exit" decision occurs only at intervals (e.g., every 4 years) rather than in a continuous fashion. Consequently, one would expect that after several decisions to reenlist, many military personnel would be committed to make the service a career. Examination of Table 2 provides further support for this rationale of a discrete versus continuous evaluation of staying in the organization. The correlations between seniority and career intention for both officers and enlisted personnel (.43 and .48, respectively) are much higher than the correlation for civilians (.17). Although the potential of more external opportunities for civilians might be a second causal factor for this lower correlation, certainly the lack of a discrete reenlistment commitment decision appears to have some merit for explaining the difference.

The data in Table 3 provide mixed support for the hypothesis that the correlational relationhips will be stronger for the military population than for civilians. Four of the six comparisons did find fairly strong differences among the military-civilian samples. Examining only the attitudinal variables (leadership and job characteristics) three out of four of these military-civilian comparisons support the hypothesis that greater relationships would exist for the military population. The ability to leave the organization at any time, the perceived potential of more external employment opportunities, and the recognition that they are non-military personnel in a military organization are possible explanations for the civilian population having weaker correlational results.

The results of this study in a military environment do support the traditional turnover models. While only slight differences were found between the results for officers and enlisted personnel, the decrease in relationships for civilians suggest that further investigation of this subgroup is warranted. Particularly, emphasis should be placed on insuring that civilians perceive the military environment as a contributing factor in their professional development. In examining the subsequent careers for social scientists in the military Clausen (1984) found that a proper professional environment was a source of superb postgraduate training and experience and resulted in establishing stimulating friendships and professional ties for the future.

References

Bass, B.M. (1981). Stogdill's handbook of leadership: A survey of theory and research. New York: Free Press.

Clausen, J.A. (1984). Research on the American soldier as a career contingency. Social Psychology Quarterly, 47, 207-213.

Finstuen, K., & Berry, G. (1981). Air Force attrition research: Analysis of pre- and post-enlistment factors. In H. Sinaiko et al. (Eds). Military personnel attrition and retention research in progress (TR-10). Washington: Smithsonian Institution.

Herman, J.B., Dunham, R.B., & Hulin, C.L. (1975). Organizational structure, demographic characteristics, and employee responses. Organizational Behavior and Human Performance, 13, 206-232.

Hom, P.W., & Hulin, C.L. (1981). A competitive test of the prediction of reenlistment by several models. Journal of Applied Psychology, 66, 23-39.

Loher, B.T., Noe, R.A., Moeller, N.L., & Fitzgerald, M.P. (1985). A meta-analysis of the relation of job characteristics to job satisfaction. Journal of Applied Psychology, 70, 280-289.

Martin. T.N. (1980). Modelling the turnover process. Journal of Management Studies, 16, 261-274.

Michaels, C.E., & Spector, P.E. (1982). Causes of employee turnover: A test of the Mobley, Griffeth, Hand, and Meglino Model. Journal of Applied Psychology, 67, 53-59.

Niebuhr, R.E. & Dansby, M.R. (1986). The relationship between leadership and job satisfaction/career commitment of Air Force personnel. Proceedings: Psychology in the Department of Defense, 269-272.

Short, L.O. (1985). The United States Air Force Organizational Assessment Package (LMDC-TR-85-2). Maxwell AFB, AL: LMDC.

Spencer, D.G., & Steers, R.M. (1981). Performance as a moderator of the job satisfaction-turnover relationship. Journal of Applied Psychology, 66, 511-514.

Steel, R.P., & Ovalle, N.K. (1984). A review and meta-analysis of research on the relationship between behavioral intentions and employee turnover. Journal of Applied Psychology, 69, 673-686.

Steers, R.M., & Mowday, R.T. (1981). A model of voluntary employee turnover. In L. Cummings & B. Staw (Eds.). Research in organization behavior (Vol. 3). Greenwich, CT: JAI.

Ward, M.P., & Tan, H.W. (1985). The retention of high-quality personnel in the U.S. armed forces. Santa Monica, CA: Rand.

Williams, L.J., & Hazer, J.T. (1986). Antecedents and consequences of satisfaction and commitment in turnover models: A reanalysis using latent variable structural equation methods. Journal of Applied Psychology, 71, 219-231.

A Behavioral Observation Scale for the Academy Training Philosophy

Major Michael A. Rosebush, Ph.D.
USAF Academy

Abstract

The United States Air Force Academy's (USAFA) standardized definition for how all supervisors assigned at USAFA are to treat their subordinates is described in a set of guidelines known as the "Academy Training Philosophy" (ATP). The researcher developed a Behavioral Observation Scale (BOS) in which subordinates could assess how often they perceived their supervisor was applying various ATP behaviors. Nine empirical tests were conducted on the BOS for its reliability and validity. Regarding reliability, the five subscales of the BOS had internal consistency coefficient alpha values ranging from .86 to .93. Additionally, the instrument's intraclass correlation on sets of four raters was .37, which proved to be higher than the intraclass correlations concurrently measured on the previously validated Job Descriptive Index (JDI). Regarding validity, a test for content validity revealed that for each of the BOS items at least 70% of the judges could accurately reallocate the item back to its original subscale. Evaluations of the instrument's concurrent validity revealed correlations of .71 and .65 respectively with the JDI's Job-in-General and Supervision scales, and .80 with a global supervisor effectiveness scale known as the Supervisor Evaluation Questions. Regarding construct validity, the test for contamination on each item revealed at least 70% of the judges believed the item satisfactorily discriminated supervisory effectiveness. Additionally, a multiple-group method factor analysis showed 44 of the instrument's 48 items loaded more heavily on the expected subscale than any of the other four subscales. Finally, there was a significant semipartial correlation between the instrument's appraisal of the supervisor and the subordinates' current military performance (after the subordinates' previous cumulative performance had been partialed out of the current semester's performance) for freshmen subordinates, but not for upperclass subordinates.

The United States Air Force Academy (USAFA) utilizes the Academy Training Philosophy (ATP), developed by Rosebush and Bryant (1984), as the model for how all supervisors assigned to USAFA are to behave when leading subordinates. The ATP model asserts that there are five key principles which need to be applied by the supervisor in order to instill in the subordinate a strong desire to excell. These five principles are:

1. Set clear EXPECTATIONS for subordinates
2. Teach subordinates the necessary SKILLS to accomplish assigned tasks
3. Provide FEEDBACK to subordinates
4. Provide appropriate CONSEQUENCES commensurate with the feedback given
5. Establish GROWTH opportunities for the subordinates

The USAFA manual <u>Applying the Academy Training Philosophy</u> (Rosebush, 1985) describes in specific detail to supervisors how to actually apply each of the five ATP principles, and this manual is taught to all cadets.

The ATP is not only the model for supervision at the USAFA, but it is also the model for the U.S. Coast Guard Academy, and is being used in various ways at Air Force Officers Training School and Squadron Officers School. Despite the widespread use of ATP in many major training centers, there had never been developed a performance appraisal rating scale for measuring how well a supervisor was applying ATP (as determined by the immediate subordinates).

While there are numerous methods for developing valid performance appraisal instruments, one of the most popular and respected is the behavioral observation scales (BOS) method. The BOS, developed by Latham and Wexley (1977) is a procedure for scale construction that attempts to measure the frequency with which certain critical behaviors are observed by the rater. An advantage of the BOS is that items are behavioral in nature, thereby describing specifically what the supervisor must do more or less of.

The purpose of the research was to develop a BOS type of performance appraisal which measured from the subordinates' perspective how well the supervisor was applying the principles of the ATP. In order to develop such a scale, the researcher went through numerous scale development steps, followed by empirical tests of the instrument's reliability and validity.

Method

The methodology used in this research involved a series of surveys given to groups of cadets from the USAFA, using field procedures.

Subjects

Five different groups of cadets were surveyed for the purpose of instrument development and validation. No cadet participated in more than one group, while all cadets were randomly selected. Group #1 consisted of 93 cadets who were very familiar with ATP, and who were used to help write the original pool of 848 possible items for the instrument. Group #2 consisted of 76 cadets who conducted both a "reallocation test" (a form of content validation) and a "test for contamination" (a form of construct validation) on the 117 non-redundant items. Group #3 consisted of 64 cadets who conducted further tests for contamination. Group #4 consisted of 243 cadet subordinates who rated their respective supervisor, using an initial version of the instrument which consisted of 65 items. Finally, Group #5 consisted of 264 cadet subordinates who rated their respective supervisor using the final 48-item instrument.

Instruments

There were a total of five, researcher-developed surveys given in the development and empirical validation of the BOS instrument. Survey #1 was given to Group #1, and required those cadets to write items which they believed were critical in defining ATP behavior. Survey #2 was given to Group #2, requiring those cadets to reallocate each item back to the original subscale from which the item was originally written. Additionally, Survey #2 required the cadets to conduct a test for contamination on each item, to determine whether the item could discriminate good from bad supervisors. Survey #3 was given to Group #3, requiring those cadets to also conduct a test for contamination. Survey #4 was given to Group #4 for the purpose of collecting data on 65 items in order to conduct an item analysis to shrink the instrument down to its final size. Survey #5 was the instrument in its final 48-item size, and was given to Group #5 for the purpose of empirical validation.

Procedure

Survey #1 was given to Group #1 in order to have cadets most familiar with ATP define the behaviors which they believed were the most critical in defining each of the five ATP subscales. This procedure produced 848 items from the 93 cadets. The researcher eliminated redundant items, thus producing 117 items. The 117 items were then tested on Group #2 by using Survey #2 to see if the items could be properly reallocated back to their original ATP subscale from which they were written. Additionally, Surveys 2 and 3 were given to Groups 2 and 3 to see if each item was free from "contamination" (i.e. capable of discriminating good from bad supervision). The results of these tests of reallocation and contamination provided the researcher with the empirical basis to reduce the instrument down to 65 items. The 65-item Survey #4 was given to Group #4, where each of the 243 cadet subordinates rated their respective supervisor. The researcher then conducted an item analysis of the results of Survey #4, and shrunk the instrument down to its final 48-item size. Survey #5 (i.e. the final 48-item instrument) was given to Group #5, and these 264 cadet subordinates rated their respective supervisor. The results of Survey #5 served as the empirical evaluation of the final instrument's reliability and validity.

Results

The final ATP instrument (i.e. Survey #5) was given to Group #5, and a total of nine separate empirical validations were conducted to assess its reliability and validity. Regarding reliability, the instrument's five subscales were measured for internal consistency by determining their coefficient alpha values. A coefficient alpha value of .80 is commonly used as a criterion for adequate internal consistency (Latham & Wexley, 1981) and all five of the ATP subscales met this criterion, ranging in values from .86 to .93.

Another reliability test was the instrument's intraclass correlation coefficient, which measured the degree of consistency between groups of raters, if these groups were to have rated the same supervisor. The intraclass correlation on the instrument proved to be .37, which was higher than the intraclass correlation of two previously validated scales for measuring supervision (i.e. two of the Job Descriptive Index scales) -- which were concurrently measured along with the ATP instrument.

Several tests of the instrument's validity were conducted. The ATP instrument was concurrently validated against two scales from the Job Descriptive Index (JDI) (Smith, Kendall, & Hulin, 1982) -- the "Job-in-General" scale and the "Supervision" scale. The JDI is the most widely used instrument in research for assessing supervisory effectiveness and is considered by many to be the most valid and reliable instrument available in the area of supervision. The ATP instrument correlated .71 with the JDI's Job-in-General scale and .65 with the Supervision scale. The ATP instrument was also concurrently validated against a third scale (known as the "Supervisor Evaluation Questions" -- a global assessment of a supervisor's overall effectiveness), and the correlation was .80.

Another validation of the ATP instrument involved conducting a semipartial correlation test to determine if a supervisor's score on the ATP instrument was significantly related to increased military performance by the subordinates (after the subordinates' previous cumulative military performance had been accounted for). The results indicated that there was not a significant relationship for all subordinates (i.e. upperclass and freshmen subordinates), but there was a significant relationship for the freshmen subordinates.

The reallocation test (i.e. a form of content validation) showed that on each item at least 70% of the judges could accurately reallocate the item back to its original subscale. The results of the test for contamination on each item revealed at least 70% of the judges believed the item satisfactorily discriminated supervisory effectiveness. A final validation of the ATP instrument involved a multiple-group method factor analysis to confirm if an item loaded most with the ATP subscale from which the item was originally derived, or if the item erroneously loaded higher on one of the other four ATP subscales. Forty-four of the 48 items loaded highest on their "correct" ATP subscale.

Discussion

The purpose of the research was to develop a reliable and valid Behavioral Observation Scale performance appraisal instrument where subordinates could assess the frequency in which their respective supervisor applied certain behaviors endorsed in the Academy Training Philosophy. The researcher had preestablished nine criteria which the ATP instrument should have been able to meet to demonstrate its reliability and validity. All nine of the criteria were met. Specifically, the instrument demonstrated reliability, as shown by its adequate internal consistency (i.e. coefficient alpha values greater than .80), and between-rater reliability (as demonstrated by having an intraclass correlation higher than either scale of the JDI, when concurrently measured).

The instrument also demonstrated content, concurrent, and construct validity. Specifically, content validity was demonstrated through use of the Critical Incident Technique (Flanagan, 1954) of item generation, plus having all items on the ATP instrument capable of being properly reallocated 70% of the time. Concurrent validity was demonstrated by having correlations of .71 with the JDI's Job-in-General scale, .65 with the JDI's Supervision scale, and .80 with the Supervisor Evaluation Questions scale. Additionally, concurrent validation was demonstrated by showing that there was a statistically significant relationship between a supervisor's score on the ATP instrument and improved military performance in the freshmen subordinates (although there was not a significant relationship when considering upperclass subordinates). Finally, the instrument demonstrated construct validity by having each item satisfy a test for contamination 70% of the time. Additionally, the multiple-group method of factor analysis confirmed that 44 of the 48 items on the instrument loaded highest on the correct ATP subscale.

These results provide extremely positive evidence for the instrument's reliability and validity. While there are some limitations which need to be considered (i.e. the degree of generalizability that this instrument has with other populations, further work needed on exploratory factor analyses, etc.), it does appear that the instrument effectively measures (from the subordinates' perspectives) how well a supervisor is applying the principles of ATP.

This research indicates that institutions which are currently using the Academy Training Philosophy (e.g. USAFA, Air Force training centers, the U.S. Coast Guard Academy, etc.) might be wise in using this ATP instrument to assess supervisor effectiveness. This instrument could also be used by institutions which are not currently using ATP, as a means of assessing whether their supervisors are applying ATP-type behaviors. If such institutions discover that their supervisors are scoring low on the ATP instrument then such data could provide impetus for developing a training program which focuses on ATP principles. There exists the potential for much more research using the ATP instrument. For instance, does the instrument demonstrate reliability and validity at other training centers where ATP is currently practiced? What about other service academies? Another intriguing area of investigation involves the generalizability of the ATP instrument to actual Air Force units or civilian units. Perhaps the instrument can effectively measure supervisory effectiveness at a wide range of settings, but only further research can answer that question.

References

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

Latham, G.P., & Wexley, K. (1977). Behavioral observation scales for performance appraisal purposes. Personnel Psychology, 30, 255-268.

Latham, G.P., & Wexley, K. (1981). Increasing productivity through performance appraisal. Reading, MA: Addison-Wesley.

Rosebush, M.A. (1985). Applying the Academy Training Philosophy. United States Air Force Academy, CO: USAFA.

Rosebush, M.A., & Bryant, C. (1984). The positive motivation model. United States Air Force Academy, CO: USAFA.

Smith, P.C., Kendall, L.M., & Hulin, C.L. (1982). The measurement of satisfaction in work and retirement: A strategy for the study of attitudes. Chicago: Rand McNally. (Original work published in 1969)

A Method for Relating Workers' Job Perceptions to Job Content Indicators[1]

Captain Rick S. Tallarigo, Ph.D.
Air Force Institute of Technology/
Civilian Institutions Program
(Bowling Green State University)

Abstract

Workers' self-reports of job characteristics such as autonomy, skill variety, task feedback, task significance, and task identity have been reliably related to self-reported worker satisfaction. As a result, a goal of many job redesign efforts has been to increase worker motivation by increasing these characteristics in jobs. Critics have charged, however, that the worker perception - worker attitude relationship is largely due to common method bias inherent in self-report measurement methods. As a result, a more definitive relationship between job perceptions and actual job design is desirable in order to (a) assess the validity of workers' perceptions of job characteristics, and (b) provide a more objective basis for job redesign interventions. This study, using canonical correlation analysis, found that workers' perceived job characteristics were significantly related to independently measured characteristics derived from job analysis.

Jobs that are enriched (i.e., have high levels of skill variety, task identity, task significance, autonomy, and feedback from the job itself) tend to be associated with desirable job attitudes and performance (Hackman & Oldham, 1975; Loher, Noe, Moeller, & Fitzgerald, 1985). Controversy has surrounded these relationships, however, because of the potential for common method bias when using self-reports of job characteristics as a basis for job redesign and as predictors of job attitudes (see e.g., Roberts & Glick, 1981). As a result, there is a need to validate job characteristics rated by incumbents. The literature suggests at least four independent criteria that can be used to validate workers' perceptions of job characteristics: (a) experimentally created job designs (typically in a laboratory setting), (b) convergence among rating sources of job characteristics (e.g., peers, incumbents, supervisors, observers), (c) significant differences between job groups in rated characteristics, and (d) results of job analyses. This last method, job analysis, would appear to be the most potentially

[1] This paper was based on the author's dissertation research under the direction of Professor Eugene F. Stone, Ph.D.

useful for providing input into the job redesign process; and is the focus of study in the present research. The research objective was to determine if relationships existed between commonly measured job characteristics (used in job redesign and worker perception-attitude-behavior theory building) and job characteristics derived from independent job analysis.

Method

Subjects

A total of 223 military and civilian workers, representing four distinct organizations and 16 job classifications, participated in the study as raters of perceived job characteristics. In addition, 34 members of the respective organizations served as job analysts. Table 1 lists the subject and job samples.

Table 1

Participants by Organization and Job Title

Organization	Job Title	Number of Analysts	Number of Workers
Food Process	Packaging Supervisor	2	2
Air National Guard	Materiel Facilities Supervisor	2	3
	Avionics Technician	3	12
	Aircraft Armament Systems Technician	2	11
	Jet Engine Technician	2	5
	Tactical Aircraft Maintenance Technician	4	19
United States Air Force	ROTC Instructor	3	13
	Recruiter	2	20
	Inventory Management: Demand Processing	2	7
	Inventory Management: Mission Capability	1	3
	Development Engineer: Electronics	2	12
	Development Engineer: Aeronautical	2	12
County Sheriff	Traffic Safety Officer	2	10
	Criminal Investigator	2	23
	Patrol Officer	2	69
Totals:	16	34	223

Measures

The researcher compiled items from Hackman and Oldham's (1974) Job Diagnostic Survey and Stone and Gueutal (1985). Based upon principal components analysis, a unit-weighted composite measure named Job Complexity was created from measures of skill variety, task identity, task significance, autonomy, feedback from the job itself, and intellectual demands. In addition, workers' perceptions of public interaction and service, dangerousness, and physical demands were also used. The job analysis technique used was the Position Analysis Questionnaire (McCormick et al., 1969). The PAQ ratings were scored to produce 13 overall job dimensions. The reliabilities of the PAQ ratings and job perception ratings were assessed prior to the canonical analysis and found to be satisfactory.

Results

The data in Table 2 represent the canonical solution relating workers' perceptions of job characteristics with the overall dimensions derived from job analysts' PAQ ratings. Three of the four canonical functions were statistically significant. The greatest characteristic root (i.e., the square of the canonical correlation between the first pair of canonical variates) and the subsequent roots were tested for significance with degrees of freedom as described in Harris (1985). The squared canonical correlations, however, indicated the amount of variance explained in the pairs of canonical variates, not the original variables. In order to interpret the substantive nature of the variates in terms of the original variables and to assess their predictive efficiency, five indicators were computed: canonical regression coefficients (Harris, 1985), structural coefficients (Meredith, 1964), variance matrix coefficients (Dunham & Kravetz, 1975), redundancy coefficients (Stewart & Love, 1968), and the squared semi-partial correlation (Darlington, 1968) between each original variable and its opposite variate. The entries in Table 2 represent the structure matrix, i.e., the correlations between original variables and the canonical variates (Meredith, 1964).

The redundancy coefficients indicated that three canonical variates of the PAQ overall dimensions predicted 60% of the variance in the perceived job characteristics; and the three variates of the perceived job characteristics accounted for 51% of the variance in the PAQ overall dimension scores.

The first canonical function clearly appeared to represent a physical demands and dangerousness dimension. Incumbents who perceived their jobs as physically demanding and dangerous also worked in jobs characterized by job analysts as requiring operation of machines and equipment, engaging in physical activities, and working in unpleasant, hazardous, and/or demanding environments.

The second root indicated that jobs rated by workers as requiring public interaction/service also were rated most highly on PAQ items associated with public, customer, and/or related contacts and low in supervising, directing, and estimating.

The third function appeared to represent job complexity. Incumbents who rated their jobs high in characteristics comprising job complexity also were found by job analysts to have regular work schedules with low amounts of task routine and repetitiveness. In this sample, workers on regular day schedules perceived their jobs as more enriched than those on atypical schedules, who performed tasks rated as more routine/repetitive by the job analysts.

Table 2
Canonical Correlation Analysis of PAQ Overall Dimensions with Perceived Job Characteristics (PJC)

Canonical function	I	II	III
Canoncial correlation	.92	.78	.51
Squared canonical correlation	.85***	.61***	.26***
PAQ Canonical Variates	PAQ1	PAQ2	PAQ3
Overall Dimensions	r	r	r
1. Decision, communicating, & general responsibility	.40	.45	-.38
2. Operating machines & equipment	.84	-.25	-.39
3. Performing clerical/related activities	.71	.38	-.45
4. Performing technical/related activities	.61	-.29	-.41
5. Performing service/related activities	.39	.26	-.55
6. Working regular day schedules	-.79	-.06	.40
7. Performing routine/repetitive activities	.67	.13	-.64
8. Being aware of work environment	.84	.19	-.40
9. Engaging in physical activities	.88	-.19	-.30
10. Supervising/directing/estimating	.16	-.67	-.33
11. Public/customer/related contacts	.28	.71	-.44
12. Unpleasant/hazardous/demanding environment	.88	-.06	-.42
13. Non-typical work schedule	.54	.20	-.66
Redundancy coefficients:	(.37)	(.08)	(.06)
PJC Variates	PJC1	PJC2	PJC3
Perceived Characteristics	r	r	r
Job complexity	.23	-.07	.78
Public interaction	-.21	.92	-.11
Physical demands	.94	-.13	-.21
Dangerousness	.92	.32	.20
Redundancy coefficients:	(.39)	(.16)	(.05)

Note: N = 223. Entries are structural coefficients.
***p < .0001

Discussion

The use of canonical correlation allowed for the relatively subjective concepts of perceived job characteristics to be linked with more objective job content indicators derived from independent job analysis. This linkage provided, in the sample studied, a validation of the workers' perceptions of their jobs using job analysis criteria. This study also provided support for a three-dimensional portrayal of perceved job characteristics similar to the Stone and Gueutal (1985) dimensions derived from multi-dimensional scaling. Finally, the methods employed in this research may be useful for job redesign efforts by providing more specific targets for consideration for redesign through the use of the PAQ items associated with the overall dimensions.

References

Darlington, R. (1968). Multiple regression in psychological research. Psychological Bulletin, 69, 161-182.

Dunham, R., & Kravetz, D. (1975). Canonical correlation analysis in a predictive system. Journal of Experimental Education, 43, 35-42.

Hackman, J., & Oldham, G. (1974). The Job Diagnostic Survey: An instrument for the diagnosis of jobs and the evaluation of job redesign projects. Office of Naval Research (TR No. 4; DTIC AD No. 779828).

Hackman, J., & Oldham, G. (1975). Development of the Job Diagnostic Survey. Journal of Applied Psychology, 60, 159-170.

Harris, R. (1985). A primer of multivariate statistics. Orlando: Academic Press.

Loher, B., Noe, R., Moeller, N., & Fitzgerald, M. (1985). A meta-analysis of the relation of job characteristics to job satisfaction. Journal of Applied Psychology, 70, 280-289.

McCormick, E., Jeanneret, P., & Mecham, R. (1969). Position Analysis Questionnaire (Form B, 8-82). West LaFayette, Indiana: Purdue Research Foundation.

Meredith, W. (1964). Canonical correlations with fallible data. Psychometrika, 29, 55-65.

Roberts, K., & Glick, W. (1981). The job characteristics approach to task design: A critical review. Journal of Applied Psychology, 66, 193-217.

Stewart, D., & Love, W. (1968). A general canonical correlation index. Psychological Bulletin, 70, 160-163.

Stone, E., and Gueutal, H. (1985). An empirical derivation of the dimensions along which characteristics of jobs are perceived. Academy of Management Journal, 28, 376-396.

Relationships between Rating Quality Measures in Laboratory-based Performance Evaluations

Rick S. Tallarigo, Ph.D.
Air Force Institute of Technology/
Civilian Institutions Program
(Bowling Green State University)

Abstract

The quality of performance ratings is commonly assessed with indices of rater errors such as halo, leniency, and restriction of range. At the group level, convergent and discriminant validity also provide evidence of rating quality. In addition to the above measures, this study utilized rating accuracy and true halo measures and examined relationships between these indicators of rating quality. It was concluded that rater "error" measures could not serve as meaningful indicators of rating inaccuracy or validity. Levels of true halo in rated performance need to be considered as well.

As their name implies, rater errors have been taken as indicators of rating inaccuracy. These errors include such measures as halo, leniency, and restriction of range. Halo has been defined as the failure to discriminate among potentially distinct dimensions of performance. Leniency can be defined as a mean rating above the scale midpoint; and range restriction is the failure to discriminate among ratees. What these traditional rater error measures have in common is a reliance on distributional assumptions. For example, both halo and range restriction assume some optimal degree of variability in ratings. In the case of halo, the smaller the variability of ratings across dimensions, the greater the halo. For range restriction, the variability is assessed across ratees: the smaller the variability, the greater the range restriction. Likewise, leniency and severity can be defined as negatively skewed and positively skewed rating distributions, respectively. None of these rater error measures depend upon a known true value for the level or range of ratings for defining rating quality. As a result, they have been described as <u>indirect</u> measures of rating quality. That is, the true levels, and distributions of performance cannot be known. Arbitrarily, then, scale midpoints and normal distributions are used as criteria against which observed ratings are compared. This rationale has been the basis of Rater Error Training, in which raters are taught to normally distribute their ratings.

An alternative approach, free of a priori distribution assumptions, is rater <u>accuracy</u>. Ratings are compared to a known true or target criterion value for the performance being evaluated. The rating accuracy approach has been used mostly in laboratory and rater training settings in which performance scenarios can be videotaped or written. The means of expert ratings are then obtained and serve as true scores.

Use of videotaped or written performance stimuli, for which true scores are known, permits examining whether or not rater error measures actually do represent rater inaccuracy, as traditional interpretation assumes. Empirical studies to date have suggested both paradoxical (positive correlations) as well as the expected negative correlations between rater error measures and rating accuracy. In addition, with true scores known for each rating dimension, the true halo for a given ratee or group of ratees can be determined. True halo can then serve as a criterion for determining the amount of halo error in the obtained observed ratings.

The present study examined the interpretability of rater error measures when compared to rater accuracy scores and convergent and discriminant validity indices.

Method

Subjects

Two-hundred and ninety-three subjects completed this study. All subjects were students enrolled either in introductory psychology or business administration. The average age of the subjects was 19 years, and approximately two-thirds were female.

Materials

Videotapes. Eight videotapes of actors delivering classroom lectures (and the true scores associated with each performance; Murphy et al., 1982) were used in the present study.

Rating Scale. Each lecturer was rated on eight performance dimensions on a scale of 1 (Very Bad) to 5 (Very Good).

Procedures

All subjects were shown a randomly selected set of four of the eight videotapes. While viewing the tapes, raters were not permitted to take notes.

Measures

The following rater error scores were computed for each rater: Observed Halo (OH), Halo Error (HE), Absolute Halo Error (AHE), Observed Leniency (OL), Leniency Error (LE), and Restriction of Range (ROR). Rating accuracy scores for each rater were computed using the formulas of Cronbach (1955) for Elevation (EL), Differential Elevation (DEL), Stereotype Accuracy (SA), and Differential Accuracy (DA). At the group level, measures of convergent and discriminant validity were computed using the multi-trait multi-rater techniques of

Kavanagh, MacKinney, and Wolins (1971). See Tallarigo (1987) for more detail on procedures and measurement methods used in the original study.

Results

Interpretation of Rater Errors

Table 1 presents the means and standard deviations for halo, leniency and true halo. A number of observations are noteworthy. First, the overall mean level of True Halo was quite high. In Pearson correlation form, the overall True Halo was approximately .95 (Fisher's z transformation = 1.83). Second, the interpretation of halo and leniency differs depending on whether true scores are considered. For example, as can be seen in Table 1, the Observed Leniency measure indicated that raters, on average, were neither lenient nor severe in their ratings, using the scale midpoint as a criterion. But when true scores were used as criteria for assessing leniency, the majority of raters were severe in their ratings, as indicated by a mean leniency error that was negative. The percentage of raters exhibiting leniency dropped from approximately one-half to one-quarter when the definition of leniency changed from using the scale midpoint to using the overall true score mean. In a similar manner, the interpretation of raters' halo tendencies changed depending on its definition. The mean Observed Halo indicated a high level of halo in the ratings (i.e., the mean interdimension correlation was .79). According to traditional interpretation, raters failed to discriminate sufficiently among the performance dimensions. But Halo Error levels (i.e., Observed minus True Halo) indicated that raters discriminated too much. That is, mean Halo Error was negative. The majority of subjects (88%) exhibited negative halo error, i.e., observed halo lower than true halo. Thus, one can arrive at differing conclusions regarding rating quality depending on whether or not true scores are considered in the definition of rater error.

Table 1.
Descriptive Statistics for Halo and Leniency

	M	SD
True Halo	1.83	.63
Observed Halo	1.07	.57
Halo Error	-.76	.70
Observed Leniency	.01	.39
Leniency Error	-.28	.39

Note: All n's = 293. Halo values are in Fisher z form.

Relationships between Rater Error and Rater Accuracy

The traditional interpretation of rater errors would presume that, as classes of variables, rater errors are negatively correlated with rating accuracy. The results in Table 2 clearly indicated that this was not the case in the present data. Except for DEL accuracy, Observed Halo was positively correlated with accuracy. The one significant relationship for Observed Leniency was positive; and for Range Restriction, one significant correlation was negative and one positive. Using Halo Error as an indicator of accuracy was also inadequate. But using the absolute value of Halo Error (i.e., Absolute Halo Error) produced more consistent results: the two significant relationships were, as expected, negatively related to accuracy.

Table 2

Correlations between Rater Error and Rater Accuracy Scores

	Rater Accuracy Scores			
Rater Error Scores	EL	DEL	SA	DA
Observed Halo	.25***	-.23***	.28***	.45***
Halo Error	.07	-.19***	.18**	.34***
Absolute Halo Error	-.07	.02	-.19***	-.35***
Observed Leniency	.63***	.06	.07	.04
Leniency Error	.80***	.06	.13	.06
Range Restriction	-.25***	.22***	-.08	.11

Note: EL = Elevation accuracy; DEL = Differential Elevation; SA = Stereotype Accuracy; DA = Differential Accuracy. $N = 293$.
** $p < .01$ *** $p < .001$

Relationships at the Group Level

There were 39 unique groups of raters in the present study. Each group was exposed to one of 14 levels of true halo in videotaped performances. This analysis examined correlations between rating quality indices and true halo, convergent, and discriminant validity after averaging the rater error and rater accuracy scores within each of the 39 groups.

Effects of True Halo. The correlations in Table 3 suggested that levels of true halo were positively correlated with convergent validity and observed halo, negatively correlated with range restriction, and unrelated to discriminant validity and rating accuracy. Also, the true halo levels did not appear to affect the rating accuracy scores.

Correlates of Convergent and Discriminant Validity. Neither the rater error nor the rater accuracy scores correlated with the validity indices in a uniform manner. However, the accuracy scores appeared more consistent than the error measures in that all of the correlations between accuracy and validity were positive. The rater error scores, however, were more mixed in their correlations with the accuracy scores.

Table 3
Group-level Correlations

	True Halo	CV	DV
True Halo	-	.56***	-.04
Rater Errors			
Observed Halo	.62****	.66****	.12
Range Restriction	- .56***	-.57****	.18
Observed Leniency	- .01	-.05	.04
Rater Accuracy			
Elevation	.31	.51***	.25
Differential Elevation	.05	.51***	.36*
Stereotype Accuracy	.27	.21	.46**
Differential Accuracy	.04	.09	.23

Note: n = 39. Signs of correlations with range restriction and accuracy scores are reversed. CV = convergent validity; DV = discriminant validity.
* $p < .05$ ** $p < .01$ *** $p < .001$ **** $p < .0001$

Discussion

This study suggested that interpretation of rating quality must consider the degree of true halo and true performance levels inherent in each ratee sample. The decision as to how much leniency, halo, or range restriction is excessive is relative to the true score distribution. Accuracy scores appeared to perform better than traditional rater error scores in the present study. Although the use of accuracy scores in operational performance evaluation settings is not practical, their use in rater training has been successfully demonstrated in other research.

References

Cronbach, L. (1955). Processes affecting scores on "understanding others" and "assumed similarity". Psychological Bulletin, 52, 177-193.

Kavanagh, M., MacKinney, A., & Wolins, L. (1971). Issues in managerial performance: Multitrait-multimethod analyses of ratings. Psychological Bulletin, 75, 34-49.

Murphy, K., Martin, C., & Garcia, M. (1982). Do Behavioral Observation Scales measure observation? Journal of Applied Psychology, 67, 562-567.

Tallarigo, R. (1987). Effects of rating delay and true halo in laboratory-based performance evaluations. Defense Technical Information Center (AD No. A194438).

Hierarchical Value Structures:
A Convergence Model of Ethical Development

William B. Carlin
Rita A. Moore
University of Colorado, Boulder

Abstract

Ethical issues are present in uncertain situations where multiple stakes, interests and values are in conflict and rules and standards are unclear. Managers continually engage in decision making that affects the lives and well-being of others. Thus, they are involved in ethical decision making. There is limited research which attempts to understand value systems in the context of the decision making process. This research is the preliminary examination of espoused values and behavioral outcomes, and proposes a new technique for the measurement of ethical development.

The interest in values, integrity, and ethical behavior is viewed by many as a recent phenomenon. This renewed popularity is due, to a large extent, to the reported abuses and scandals of businesses. The stories of lying, bribes, false reports and payoffs are just a sample of the adverse publicity on which the public judges organizations. Because the public has given business a low grade in honesty and trustworthiness, society is demanding that businesses and individuals address the issue of ethics. Society is demanding that business look farther than just economic issues and that they be responsive to others besides stockholders. This area of responsibility is also receiving attention from the U.S. Congress by holding organizations accountable for the ethical behavior of their managers. This concern is not limited to for-profit organizations but has also focused on the public sector. Honor incidents at the U.S. Military Academy and the U.S. Air force Academy resulted in both organizations instituting several new policies and programs dealing with honor and ethics. Society is not limiting its focus to organizations but is also demanding that its public figures display ethical behavior witnessed by the withdrawal from candidacy of several individuals seeking political office. Society is clamoring for an answer to these issues as evidenced by grants and endowments to universities to address this area. There is a dichotomy that exists within society in response to these issues. On the one hand, society seems to be clamoring for ethics to be taught, and yet parents are not willing to let their children be taught "ethics" in the schools. By the time a person reaches the undergraduate or graduate level of school, his/her ethical values are already in place, typically without any formal guidance or study.

Annually, over 12,00 ethics courses are offered by undergraduate business schools (Murray, 1987). The purpose of these classes is to force students to think through all the ramifications of an issue and to decide for themselves what is right and what is wrong, as well as to think and look at problems in new ways (Lee,1986; Murray,1987). A question raised by many is the effectiveness of teaching ethics in the classroom. While Hoff (1989) suggests that individuals can learn to make ethical decisions and, in fact, change their ethics, Lee (1986) contends that while ethical theories and cases can be taught, classes cannot teach someone to be ethical and this method is doomed to fail.

The military service academies are perhaps the most visible organizations to make inroads in teaching ethics and moral behavior in an educational environment. The most obvious symbol is the implementation of an honor code, by which all cadets must vow to abide. By having an honor code, the service academies are able to create an environment where honorable behavior can be learned and reinforced. "True honor requires active support of positive principles rather than simple abstinence from wrongdoing." (Honor Code Reference Handbook of the Air Force Cadet Wing, 1987) Research on ethical behavior of business graduate students supports this idea of reinforcement. Hagerty & Sims (1987) conclude that the threat of punishment may have a restricting impact on unethical behavior. They suggest that organizations must be willing to specify policies and guidelines for maintaining ethical behavior. However, they caution that merely developing a Code of Conduct will not guarantee its practice. If there are no sanctions associated with unethical behavior, an individual who is personally not endowed with a strong value system will most likely succumb to temptation.

A Convergence Model of Ethical Development

How does one get an individual or an organization to utilize their hierarchy of values to consider the long term consequences of its actions rather than immediate gains? Research in value systems has tended to be seen as a function of organizational culture which is in turn related to organizational climate and effectiveness. Where does an individual get the principles or standards that they reflect on in determining what to do? Moral behavior has been approached from either the moral development perspective of Lawrence Kohlberg (1976) or an espoused values perspective such as Milton Rokeach (1973). Recent studies (e.g., Bridges & Priest, 1983; Priest, 1987; and Roffey, Bryant, & Porter, 1988) have failed to support a relationship between advanced stages of moral development and predictability of desired behaviors. There is little empirical research which attempts to understand value systems in the context of the decision making process. The influence of an individual's values on ethical behavior was addressed in an experiment conducted by Hegarty & Sims (1978). They concluded that even with the strongest reinforcement for unethical behavior, some individuals were "perfectly" ethical. Scholars agree that an individual approaches an ethical dilemma with a preconditioned set of values which determines their behavior.

This preliminary examination of espoused values and behavioral outcomes proposes a new technique for the measurement of ethical development. Four populations are compared according to a revised Rokeach Value Survey. Samples are drawn from populations of business students at two secular universities, a sectarian college, and a military academy. Two series of propositions are set forth. The first associates socially undesirable or deviant behavior with the failure to discriminate between instrumental and final values. Failure to discriminate results from developmental lags in socialization and adaptive responses to the demands of the current and past social environment. The second set of propositions argue that individually espoused values are a better predictor of ethical behavior than the mechanics of the decision process alone. The success of efforts to instill ethical behavior in individuals is best measured by convergence over time of the value structures of the individual and those of a model group (e.g., faculty and military training staff) which exhibits the desired behaviors. Hypotheses tested include:

H1: Deviant populations will tend to rank the instrumental values higher than final values.

H2: Non-deviant populations will rank the final values higher than the instrumental values.

H3: Persons educated in a sectarian and/or liberal arts environment will rank the more abstract terminal values above the more concrete terminal values.

H4: Persons educated in a secular and/or military environment will rank the more concrete terminal values above the more abstract.

H5: There will be a convergence over time of value systems

a: within the student group.

b: within the model group.

c: between the student group and model group.

d: between the student group and the model group, which will be accelerated at the time of initial contact and during times of crisis and exposure to materials highly value concentrated.

Method

The survey instrument utilized in this research is a modification of the Rokeach Value Survey (RVS) (Rokeach,1973) which is a compilation of the most commonly held values. Customary usage of the RVS involves the creation of two separate hierarchies of values, one for the terminal values and one for the instrumental values. The instrument used for this study combines the elements of the two hierarchies alphabetically into one listing of values to be placed in rank order by the respondents. This modified instrument ranks values both within-value-group and between-value-group. Mean rankings are then examined across the various sample populations. This instrument achieves a reliability coefficient of .96. Demographic information regarding age, education, ethnicity, etc. were also collected.

Implications

Initial analysis of the results leads us to suggest several implications. 1) While ethical theory classes are necessary they are not sufficient. 2) Ethics education can not be limited to a single dose method. 3) Value systems of faculty, staff, and the institution itself are more powerful a teaching tool than the traditional classroom approach. 4) The convergence model has implications for the predictability of ethical decisions versus prediction of deviant behavior. The linkages established by this preliminary examination of expressed value systems and behavioral outcomes will have far reaching implications for the inclusion of value-based variables into the decision making process and the teaching of ethics and moral behavior.

References

Bridges, C. F. & Priest, R.P. (1983). Development of values and moral judgements of West Point cadets (Report No. USMA-OIR-83-002). West Point, NY: Office of Institutional Research, United States Military Academy.

Hegarty, W. & Sims, H. Jr. (1978). Some determinants of unethical decision behavior: An experiment. Journal of Applied Psychology, 63, 451-457.

Hoff, J.W. (1989). Students, ethics and surveys. Journal of Business Ethics, 8, 823-825.

Honor Code Reference Handbook of the Air Force Cadet Wing. (1987). United States Air Force Academy, CO.

Kohlberg, L. (1976). Moral stages and moralization, the cognitive-development approach. In T.Lickona (Ed.), Moral development and behavior. New York: Holt, Rinehart & Winston.

Lee, C. (1986, March). Ethics training: Facing the tough questions. Training, pp. 30-41.

Murray, T. (1987). Can business schools teach ethics. Business Month, 1, 24-26.

Priest, R. F. (1987). Feasibility of a predictor for vulnerability to dishonesty (Report No. USMA-OIR-87-004). West Point, NY: Office of Institutional Research, United States Military Academy.

Roffey, A.E., Bryant, C., & Porter, D.B. (1988). Moral development of officer candidates in the honor code environment of the United States Air Force Academy. Unpublished manuscript, United States Air Force Academy, Dept of Behavioral Sciences and Leadership.

Rokeach, M. (1973). The nature of human values. New York: Free Press.

Transcendentals: An Overlooked Factor in Leadership Research and Training

Russell Doll, Ph.D.

University of Missouri-Kansas City

Abstract

Leadership research and training deals with quantifiable behaviors. These behaviors may be appropriate for formal, static situations, but inappropriate for fluid situations. An internalized foundation of transcendental beliefs, not measured by quantifying, may provide inner-strengths appropriate to fluid situations.

Since the mid-1950's most of the leadership research has been of three types, with some minor variations. One type has concentrated on quantitative analysis, trying to isolate characteristics of leaders so as to increase productivity in certain tasks and identify those who had these characteristics. This kind of research was interested in identifying straightforward leadership behaviors with little attention paid to social dynamics. Another type attempted to analyze the effects of group dynamics on leaders and the reciprocal relations between leaders and groups. While quantification was involved, the emphasis was on socio-psychological factors. Attempts were also made to develop broad theoretical frameworks and construct "laws" governing groups and leaders. Homans' book, The Human Group, is an example of this kind of synthesis in theory building and identification of "laws." The work of Cartwright, Strodbeck and Katz is typical of the research done at that time. My third example is a recent development and includes elements of both the socio-psychological and the purely quantitative. It is called transformational leadership, the concept based on research which attempts to recognize non-quantifiable factors, those elusive factors of charisma and leadership style. The watershed article dealing with this research appeared in 1971 in Management Science followed by, among others, books such as Beyond Rational Management, Leadership Challenge and Managing Excellence.

I now want to discuss not a distinct fourth type of research but whether the possibility exists that there may be overlooked factors which govern leadership behaviors examined in the former three types. This factor has to do with internalized belief systems which influence leadership on both a conscious and unconscious level, in predictable and unpredictable situations. The factor is one related more to philosophy, and to Platonic and Neo-Platonic philosophy in particular. It is related to the belief in transcendentals and the acceptance of transcendental ideas being as real as what we can quantify or observe.

Unlike the former types of research, which are interested primarily in quantifiable behaviors and productivity, transcendentals are not amenable to quantification and not concerned with productivity. They are internalized ideas which shape behavior for other than mainly productive purposes although certain kinds of productivity result. Transcendentals, I would argue, serve to strengthen, emotionally and spiritually, formal leaders in situations of chaos and danger and they tend to lead to altruistic behaviors. Transcendentals also serve to create new leaders as the new leader draws from the inner-strength provided by the transcendental. Transcendentals may serve as the determining factor in maintaining a leaders strength and determination when normative structures collapse.

Transcendentals are often ignored in research and in academia since they do not fit into todays reality, being considered irrelevant holdovers from a past way of thinking; superfluous and quaint.

I want to briefly discuss why we ignore transcendentals and in so doing provide a clearer idea of what I mean by the term. At the end of the 13th century a debate took place critical to determining the direction of Western thought. Neo-platonists argued a reality of the intellect. Ideas, or transcendentals, such as Truth, Beauty, Self-Sacrifice, Honor and Freedom were as real as what we can see or measure. Existing on a higher level these transcendentals would guide our actions, providing "purer" behaviors which would reflect the transcendental realities. Opponents, however, lead by William of Occum, argued that reality was only what could be measured and quantified. Transcendentals were superfluous when examining real life. The nominalists won the debate. The neo-Platonists were expelled from the universities. Transcendentals disappeared from the curriculum as an element of reality, and disappeared as a part of the mind of Western man. This new mind which now recognized only the senses as real, marched towards the triumph of science in ordering the senses and examining situations. Quantification was governed only by reality as reality was determined by man. Almost all aspects of our being became quantifiable as being and becoming, morality and ethics, were defined as man chose to define them. Consequently, our research in leadership ended up being segmented by situations, particularistic in discussions, task specific in applications, cognitive/pragmatic in training, and primarily self examining and quantitative in design.

Now the emphasis on quantification in identifying leaders is not without its strengths. Such an approach is valuable in identifying "specific" leadership characteristics which predict "success" within a variety of job situations, provided these situations are static in outcome and are governed by accepted norms. The quantitative and productive research approach also assists in predicting leadership characteristics, thereby, allowing for screening and enhancing productivity. So, quantitative and productive research is appropriate for predicting traits and training leaders for situations which are formal in structure or predictable in sequence and task.

But this approach may be inappropriate for predicting and training for leadership when situations are fluid and when all familiar and recognizable norms have broken down, have been replaced, or are reforming. In these situations, I would argue, transcendentals provide strengths, while many of the quantitatively or productivity assessed measures no longer are appropriate to the situation. These transcending factors lay dormant within an internalized belief structure. But at certain times they provide a foundation of strength and meaning from which a person acts.

By our ignoring a "foundation of belief", which transcends leadership defined as characteristics, productivity or task completion, we severely limit the identification, of and training for, leadership by "vision" or through "inner strengths." We fail to recognize and train for leadership in periods of long term deprivation or when formal and predictable situations no longer exist. Transcendentals may be the factor which explains why some people rise to meet moments of moral truth while others shrink back.

For example, could anyone have predicted the leadership which surfaced among the American soldiers after the surrender of Batan and Corregidor? Many from the ranks who never showed leadership capabilities before, and many in peripheral roles, took over while many in the formal roles stood aside. Similar situations developed amongst prisoners of the North Vietnamese.

The actions of the new leaders were selfless and altruistic, self chosen in the face of quantitative evidence of danger to ones self.

And, why did some ordinary citizens in a small French town risk their lives to harbor Jews? And why did they assume leadership in persuading their neighbors to act similarly?

In a recent book, "The Altruistic Personality: Rescuers of Jews in Nazi Europe," the authors present a theory of leadership which bears on my discussion. Controlling for many factors they found one main factor which distinguished rescuers from non-rescuers. "What most distinguished (rescuers were) embedded relationships." The most "embedded relationship" was the family and a family with deep religious commitment to caring about others. A Dutch rescuer said, "It's not because I have an altruistic personality. It's because I am an obedient Christian. The Lord wanted us to rescue those people and we did it." How much more of a Transcendental element in leadership is there? And more importantly, these findings seems to run counter to Adorno's claims that religious "types" were conservative personalities, inflexible and tending towards fascism; a claim incorporated in many quantitative judgements on leadership potential.

I think, today, even William of Occum would raise the caution that our almost totally quantitative minds, geared towards productivity, could be overlooking a transcendental factor in leadership. We might even be misinterpreting some data, as in Adorno's case, by concluding that those with internalized transcendentals are inflexible in leadership positions. And, finally, by denying the existence of the Platonic reality of the intellect in favor of the nominalist reality of the senses alone, might we, in our training of the young (note, I did not say education) and of the cadets, be missing a critical factor which might, in other situations, sustain and strengthen them for a trial of self or a moment of moral truth.

References

Adorno, T. (1950). The Authoritarian Personality. New York: Harper and Row.

Bradford, D. and Kuhn, A. (1984). Managing Excellence: New York: Wiley

Homans, G. (1950). The Human Group. New York: Harcourt Brace.

Johnson, P. (1983). Modern Times. New York: Harper and Row.

Konzes, J. M. and Posner, B. (1987). Leadership Challenges: San Francisco: Jossey Bass

Mullen, J. D. (1981). Kierkegaard's Philosophy: Self-Deception and Cowardice in the Present Age. New York: Mentor Books

Oliner, Samuel and Pearl. (1988). The Altruistic Personality: Rescuers of Jews in Nazi Europe. London: Collier/McMillan.

Pelikan, J. (1984). The Vindication of Tradition. New York: Yale University Press.

Quinn, R. E. (1988). Beyond Rational Management. San Francisco: Jossey Bass.

Schueffer, F. A. (1976). How Should We Then Live? The Rise and Decline of Western Culture: Old Tappan, New Jersey: F. H. Revell Co.

Weaver, R. (1988). Ideas Have Consequences. Chicago: University of Chicago Press.

Field Validation of the Military Equal Opportunity Climate Survey (MEOCS)[1]

LtCol Mickey R. Dansby, Ph.D.
Defense Equal Opportunity Management Institute
Dan Landis, Ph.D.
University of Mississippi

Abstract

The present study reports field validation results for the Military Equal Opportunity Climate Survey (MEOCS). A preliminary version of MEOCS (Landis, Fisher, & Dansby, 1988) was modified and validated using a sample of 1650 military personnel representing all military Services and key racial/ethnic/gender groups. Data were gathered at six sites reflecting geographic and mission heterogeneity. Five factors were identified using factor analysis: Sexual Harassment, Differential Command Behaviors, Positive Equal Opportunity (EO) Behaviors, Overt Racist/Sexist Behaviors, and "Reverse" Discrimination. Factor reliabilities ranged from .50 to .90. Predicted factor score differences between racial/ethnic/ gender groups were interpreted as validating MEOCS. Additional comparisons based on subscales from other instruments included in the survey package (modified Racial Attitudes and Perceptions Survey; job satisfaction and commitment items; perceived work group effectiveness) further validated MEOCS. MEOCS will be available for command use in assessing EO climate.

A significant amount of research has been conducted on organizational effectiveness, organizational climate, and related constructs. However, relatively little research has been directed toward equal opportunity (EO) climate within organizations. Some have pointed to the impact of civil liberties climate on organizational outcomes (Scheinfeld & Zalkind, 1987) and the effect of organizational climate on EO and affirmative action (Sargent, 1978), but the personal and organizational influences of EO climate remain largely unexplored.

Landis, Fisher, and Dansby (1988) reported the construction and preliminary validation of an EO climate assessment instrument designed for use in military organizations. This instrument, now known as the Military Equal Opportunity Climate Survey (MEOCS), is based on a view of EO climate which emphasizes expectations by individuals within an organization that opportunities, responsibilities, and rewards will be apportioned based on individual merit rather than race, gender, religious, or other irrelevant factors. Individual perceptions of EO climate, whether valid or not, are perceived as influencing responses to the military environment, job performance, commitment to the military, and, ultimately, unit effectiveness (Landis, Fisher, & Dansby, 1988).

[1]This paper reflects the opinions of the authors only and should not be construed as official policy of the Department of Defense or any organization.

The impact of EO climate may be extensive in organizations. At the very least, an "atmosphere of discrimination" serves as a basis for legal action (under Title VII of the Civil Rights Act of 1964) by individuals against the organization (Baxter, 1985; Laurent, 1987). Bowers (1975) found a negative relationship between organizational climate and felt discrimination; however, work by Parker (1974) and Pecorella (1975) suggests the relationship may be keyed to interpersonal interactions within particular work groups. Griesemer (1980) found significant correlations between racial climate and unit effectiveness.

Other researchers have demonstrated consistent differences between racial, gender, and officer/enlisted groups in perceptions of EO and organizational climate. Brown, Nordlie, and Thomas (1977) found significant differences between whites and blacks in how they viewed the "race problem" in the Army. Spicher (1980) demonstrated that Air Force men perceived a significantly more favorable organizational climate than military women; similarly, officers perceived the climate more favorably than enlisted members. A survey conducted by the Army also showed differences between minorities and whites and between enlisted members and officers on items dealing with EO (Soldiers Report IV, 1986).

Reported differences in perceptions of men and women, minorities and whites, and officers and enlisted members should obtain in a valid measurement of EO climate. The purpose of the present study was to further develop the factor structure of MEOCS and to validate the instrument based on these predictions.

Method

Five military sites were initially selected for the validation study, based on the following criteria: all military Services (including the Coast Guard) must be represented; each site must have representative numbers of female and minority members; the total group must reflect a variety of missions and geographic locations; one site must be overseas. An additional site was added from the Air Force because of its convenience, wide representation of military women, and suitability for testing administration procedures. The sites were not selected to be representative of the Services, but to provide locations where a wide demographic range of military personnel could be accessed directly in on-site survey administrations and interviews.

Military members at each site were selected according to a purposive stratified sample reflecting racial/ethnic, gender, and officer or enlisted categories. Among the 1650 respondents, Service representation ranged from 97 (Coast Guard) to 607 (Air Force). All major combinations (i.e., black/white, male/female, and officer/enlisted) were represented by at least 200 respondents, with the exceptions of white female officers (109), black female officers (37), and black male officers (80). The lower incidence of these groups was due to their sparse representation in the Services (white women comprise 11.5% of military officers; black women, 1.5%; and black men, 5.1% [Source: Defense Manpower Data Center, June 1989]).

A team of researchers reflecting racial/ethnic and gender diversity administered the 157-item survey on site in a group setting. Respondents rated 88 MEOCS items on a scale of 1 to 5, according to their estimation of the likelihood listed behaviors (critical incidents) may have occurred at their duty location during the last 30 days. The survey package also included 12 items measuring commitment to the Service, 5 job satisfaction items, 6 items assessing perceived work group effectiveness (sources for these items are described in Landis, Fisher, & Dansby, 1988), 27 items adapted from the Racial Attitudes and Perceptions Survey (RAPS; Hiet et al., 1978), several demographic items, and global items asking whether the respondents perceived they had been victims of discrimination. After each administration a random sample of the group was asked to remain and respond to a structured interview concerning the readability of the survey and its perceived validity for the stated purpose. Questionnaires and computer-scorable answer sheets were collected by the researchers after each administration.

Results

The MEOCS portion of the questionnaire was factor analyzed using principal component analysis with varimax rotation. Five primary factors were identified, accounting for a total of 83.2% of the variance. The factors are listed in Table 1, along with their psychometric properties.

Table 1
MEOCS Factor Structure

Factor	No. of Items	Eigenvalue	% of Var.	Alpha
I. Sexual Harassment	21	20.93	58.0	.93
II. Differential Command Behaviors	11	4.85	13.0	.90
III. Positive EO Behaviors	8	1.85	5.0	.77
IV. Overt Racist/Sexist Behaviors	6	1.54	4.3	.68
V. "Reverse" Discrimination	4	1.03	2.9	.50

Table 2 presents a summary of significant factor score differences between various racial/ethnic, gender, and personnel category groups.

Commitment to the Service scores were higher for whites than blacks ($F[1,1623] = 10.78$, $p < .01$); for males than females ($F[1,1623] = 5.01$, $p < .05$); and for officers than enlisted members ($F[1,1623] = 20.92$, $p < .0001$). Officers scored higher than enlisted members on job satisfaction ($F[1,1623] = 15.85$, $p < .0001$). Whites rated their work groups higher in effectiveness than did blacks ($F[1,1623] = 10.56$, $p < .01$); officers rated work group effectiveness higher than enlisted members did ($F[1,1623] = 22.88$, $p < .0001$).

On the RAPS portion, blacks perceived more discrimination against minorities than did whites ($F[1,1623] = 369.99$, $p < .0001$); women more than men ($F[1,1623] = 56.26$, $p < .0001$); and black officers than black enlisted, while the reverse was true for whites ($F[1,1623] = 12.80$, $p < .05$). On the RAPS "Reverse" Discrimination factor, males perceived greater occurrence than did females ($F[1,1623] = 10.90$, $p < .001$); whites perceived more than blacks ($F[1,1623] = 45.85$, $p < .0001$); and black officers perceived more than black enlisted, while the reverse was true for whites ($F[1,1623] = 8.99$, $p < .01$).

On a third RAPS factor, males agreed more than females that the races should be kept separate ($F[1,1623] = 10.99$, $p < .001$), and blacks agreed more than whites that the races should be separated ($F[1,1623] = 10.20$, $p < .01$).

Table 2
Significant MEOCS Factor Score Differences

Factor	Significant Differences
I	Women perceived more sexual harassment behaviors occur than did men ($F[1,1623] = 39.62$, $p < .0001$); blacks perceived greater occurrence than whites ($F[1,1623] = 28.39$, $p < .0001$); black officers perceived greater occurrence than black enlisted, but white enlisted perceived greater occurrence than white officers ($F[1,1623] = 8.19$, $p < .01$).
II	Women perceived greater occurrence of differential command EO behaviors than did men ($F[1,1623] = 21.54$, $p < .0001$); blacks perceived greater occurrence than whites ($F[1,1623] = 241.78$, $p < .0001$); black officers perceived greater occurrence than black enlisted, but white enlisted perceived greater occurrence than white officers ($F[1,1623] = 5.39$, $p < .05$).
III	Whites perceived greater occurrence of positive EO behaviors than did blacks ($F[1,1623] = 72.37$, $p < .0001$); officers perceived greater occurrence than enlisted ($F[1,1623] = 21.11$, $p < .0001$).
IV	Blacks perceived greater occurrence of racist/sexist behaviors than did whites ($F[1,1623] = 13.81$, $p < .001$); white enlisted perceived greater occurrence than officers ($F[1,1623] = 5.17$, $p < .05$).
V	Males perceived greater occurrence of "reverse" discrimination behaviors than did females ($F[1,1623] = 6.45$, $p < .01$).

Discussion

The results are interpreted to support the use of MEOCS as a measure of EO climate. The psychometric properties are generally considered acceptable; however, additional items will be generated in an attempt to improve the reliability of Factor V, "Reverse" Discrimination, because of the relatively low alpha value (.50). The general pattern of results shows the predicted agreement between MEOCS and other instruments. As predicted, there were significant differences in how males and females, minorities and whites, and officers and enlisted members viewed discriminatory behaviors. Comparisons between RAPS results and MEOCS results also support the construct validity of MEOCS.

The MEOCS is undergoing final revisions to reduce the number of items and improve the reliability of Factor V. The survey will be available for administration within military commands on a by-request basis. Plans are being developed for administering MEOCS to a normative probability sample from each Service. Finally, MEOCS is being used to assess the impact of EO climate on mission effectiveness, as indicated in Landis, Fisher, & Dansby (1988).

References

Baxter, R. H. (1985). Sexual harassment in the workplace. New York: Executive Enterprises Publications.

Bowers, D. G. (1975). Navy manpower: Values, practices, and human resources requirements. Ann Arbor, MI: University of Michigan, Institute for Social Research.

Brown, D. K., Nordlie, P. G., & Thomas, J. A. (1977). Changes in black and white perceptions of the Army's race relations/equal opportunity programs--1972 to 1974. Technical Report #TR-77-B3. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Griesemer, H. A. (1980). Racial harmony, leadership, and unit effectiveness in combat units: An exploratory assessment of causal relationships. Research note 84-71. Monterey, CA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Hiett, R. L., McBride, R. S., Fiman, B. G., Thomas, J. A., O'Mara, F. E., & Sevilla, E. (1978). The Racial Attitudes and Perceptions Survey. Technical paper 338. Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Landis, D., Fisher, G., & Dansby, M. R. (1988). Construction and preliminary validation of an equal opportunity climate assessment instrument. Proceedings, Psychology in the Department of Defense, Eleventh Symposium. Colorado Springs, CO: U.S. Air Force Academy.

Laurent, A. (1987, June 8). Hostile environment gives twist to sex bias case. Navy Times.

Parker, W. S. (1974). Differences in organizational practices and preferences in the Navy by race. Ann Arbor, MI: University of Michigan, Institute for Social Research.

Pecorella, P. A. (1975). Predictors of race discrimination in the Navy. Ann Arbor, MI: University of Michigan, Institute for Social Research.

Sargent, A. G. (1978). Affirmative action: Total systems change. University of Michigan Business Review, 30, 18-26.

Scheinfeld, D., & Zalkind, S. S. (1987). Does civil liberties climate in organizations correlate with job satisfaction and work alienation? Psychological Reports, 60, 467-477.

Soldiers Report IV. (1986). Department of the Army, Directorate of Human Resources Development.

Spicher, C. R. (1980). Equal opportunity and treatment: Perceptions of United States Air Force military men and women. Maxwell Air Force Base, AL: Air Command and Staff College.

The Possibility of American Military Women Becoming Prisoners of War: A Challenge for Behavioral Scientists

Major Wayne E. Dillingham*
Assistant Professor of Law
United States Air Force Academy

Abstract

Regarding women in combat, the "DoD Risk Rule" is an official Department of Defense policy pronouncement which states that the risk of capture is among the proper criteria for closing certain positions or units to women. Many argue that this risk is a proper consideration because, they believe, American military women serving as prisoners of war would have substantial adverse consequences for national security. Such arguments are based on speculation. However, behavioral scientists could—and should—conduct research to reduce the level of speculation involved. These studies would assist in either supporting or refuting such arguments and, in so doing, would provide an invaluable service to our nation's policymakers.

Events associated with the United States' recent invasion of Panama demonstrate that, in our future armed conflicts, there is an ever increasing likelihood that American military women will be directly engaged in front-line combat operations. Nevertheless, there remains a body of federal law and, more importantly, policy which purports to eliminate or at least reduce the risk that these women will be exposed to direct combat, hostile fire, or capture. However, neither the law nor policy is chiseled in stone. America's policymakers are free to change them at any time. In fact, many in Congress and throughout our federal government are now taking a fresh look at the issue of women in combat. Behavioral scientists are in an ideal position to provide them assistance—especially as these policymakers consider the possibility that American military women may be captured and become prisoners of war.

---

*The views expressed herein are solely those of the author and do not necessarily reflect those of either the United States Air Force Academy, the United States Air Force, or the Department of Defense.

The Current Policy

The Defense Department's current policy regarding women in combat is set forth in the so-called "DoD Risk Rule." It is an official policy pronouncement which, in effect, is couched in terms of a balancing test based on the type, degree, and duration of certain risks—one of which is the risk of capture.

Is the risk of capture a valid consideration in this regard? Is it properly among the criteria to be considered in closing certain positions or units to women? Or, should we remove risk of capture from the Risk Rule's balancing test? In so doing, are we Americans prepared to say that we are no more concerned about our military women becoming prisoners of war than we are our military men?

From an ethical perspective, I believe that we should treat our military women with dignity—as rational human beings with a free will, informed by reason. If these women choose to expose themselves to the risk of capture and to the possibility of cruel and inhuman treatment or even death, then we should not deny them that freedom of choice.

Many, on the other hand, advocate a utilitarian approach. They contend that American military women serving as prisoners of war would result in substantial adverse consequences for our national security. However, they cannot demonstrate that such adverse consequences would actually result. Their arguments are based on speculation. For all they know, women in the prisoner of war environment may even have beneficial effects. Such is the nature of speculation. The results could go either way.

Behavioral scientists could help in resolving many of these issues. They are ideally suited to conduct the necessary research and provide the proper analyses so as to reduce the level of speculation involved. Such studies would assist in either supporting or refuting the above arguments. Furthermore, through these studies, behavioral scientists would provide an invaluable service to our nation's policymakers regarding one aspect of this complex, controversial issue we call "women in combat."

The Challenge

No amount of research or study can resolve all of these issues or remove all of the speculation regarding the consequences of American military women as prisoners of war. Only the experience of American combatant women in actual captivity could answer all of the questions. However, behavioral science research could provide substantial insight and assistance as we attempt to predict those consequences. Moreover, because our policymakers are reevaluating our position, we need the research now. After all, we do want their decisions to be as well informed as possible.

If American military women should ever become prisoners of war in any future conflict, it will not be a historical first. During World War II, for example, eleven Navy nurses and sixty-six Army nurses were captured by the Japanese in the Philippines and held prisoner for thirty-seven months. Although the experience of these women is important for our analysis, we should not draw any final conclusions from it. Modern captors, in their interrogation and exploitation attempts, generally subject their prisoners to treatment that is substantially more cruel and inhuman than that experienced by World War II prisoners. Furthermore, female combatants who, for example, may have just bombed an enemy's country are not likely to be viewed in quite the same manner—nor receive the same treatment—as the female nurses whom the enemy just captured in a field hospital.

For the sake of analysis, we should begin by assuming that American women who become prisoners of war in future conflicts will be subjected to the most extreme physical and psychological torture imaginable as the enemy attempts to interrogate and exploit them. We should also assume that the enemy's techniques regarding women will include rape, threats of rape, sexual assault, and other forms of sexual misconduct. Given this framework for analysis, behavioral scientists should begin their research.

In this framework, the question of pregnancy almost invariably arises. Therefore, for the sake of a complete discussion, it should be addressed. However, from a medical perspective, it must be emphasized that such pregnancies are most unlikely. Given the extreme stress and poor diet generally associated with the prisoner of war environment, most if not all of these women will likely experience amenorrhea (absence of the menses) and, therefore, the likelihood of pregnancies will be decreased.

Amenorrhea is a symptom which may result from any one of several causes—one of which is hypothalamic dysfunction. Although hypothalamic amenorrhea is not yet fully understood, it is frequently diagnosed among women who are "serious athletes, ballet dancers, women under severe stress, and those with sudden, large losses of body fat" (Burke & Lin, 1988). Amenorrhea is also reported to be common among new female cadets at the service academies. In fact, the Nurse Practitioner in the Air Force Academy's Cadet Clinic Primary Care has informed me that roughly 80 percent of the female cadets who enter the Academy as freshmen do not experience a menstrual period from the time of arrival in the summer until returning to their respective homes for the Christmas holidays (Guzman, 1989). In light of all of the above observations and experiences, we can expect that, if any American women do experience pregnancy during captivity, it will probably be because they conceived before capture or were raped shortly after capture. The chance that they could become pregnant at any subsequent time is substantially reduced—even if the captivity should extend for a period of years.

Even if pregnancy is only a remote possibility, we must still consider that possibility. It, then, raises numerous questions for behavioral scientists. For example, how well would the women be treated during pregnancy? What are the chances of miscarriage? What effect would such a miscarriage have on the prisoner of war? Should the United States have an official policy position concerning women who desire abortions under such circumstances? Would such abortions create dissension among fellow prisoners of war? Would these abortions be performed by enemy personnel? By fellow prisoners? Or, would they have to be self-induced? If we expect American military women to become prisoners of war someday, then we should begin addressing these questions today.

There are numerous other questions which behavioral scientists should likewise address. At the very least, they should ask: How would American military women perform in captivity? Are there physical or psychological differences which would prevent them from performing as well as their male counterparts? If so, are there likely to be any adverse consequences for national security? Or, on the other hand, are the women likely to perform better than their male counterparts? Furthermore, what would be the likely effects of American military women being present in the prisoner of war environment? Would their presence have a detrimental or a beneficial effect on the male prisoners? What types of relationships would likely form and what would be the effects of these relationships? Would male prisoners have a tendency to be overly protective? What new leadership challenges would the presence of women present—both in the prisoner compound and on the home front? What would be the effect on public support for the war—i.e., the national will—when the American public realizes that our military women are being held captive? Would we as a society be capable of accepting that our women are being physically and mentally tortured and abused in the cruelest and most inhuman ways? Could we cope with that fact? Again, what would be the likely consequences for national security?

Although it would by no means dispose of any of these issues, I suggest that behavioral scientists at least begin their research by observing the Code of Conduct training in our services' level-C SERE [Survival, Evasion, Resistance, and Escape] Training Programs. These high-stress training programs include simulated captivity in mock prisoner of war compounds and offer excellent opportunities for observation, data collection, and analysis. It would, of course, be a mistake to conclude that the performance of women in a training environment necessarily indicates how they would perform in actual captivity. However, it should provide at least some insight—and it is a source of relevant information that, as of now, is being overlooked.

The United States Air Force level-C SERE Training Program located at the Air Force Academy provides one example in this regard. Each summer, the Air Force conducts three 3-week periods of such Code of Conduct training at the Academy. Although no detailed records of student performance are retained after each summer, my personal observations as one of the Officers-in-Charge of the Resistance Training Lab (i.e., prisoner of war compound) and my discussions with members of the permanent SERE staff indicate that the female students are generally outstanding performers and are very adept at resistance techniques—perhaps even better than their male counterparts. More specifically, during the three SERE training periods of 1989, female students distinguished themselves by winning the following performance awards: (1) during 1st period, one "Best Student in Squadron" award—a squadron consisting of 100-120 male and female students, (2) during 2nd period, one "Best Student in Squadron" award, and (3) during 3rd period, two "Best Student in Squadron" awards, one of whom went on to win the "Best Overall Student" award—in other words, the best among 450-480 male and female students. Looking back over the experience of recent years, the SERE staff confirm that these 1989 awards are representative of the numbers and types of awards presented to their female students in previous years. Nevertheless, there have been no records kept, no studies performed, no analyses made, nor any such recommendations offered as to how we could improve this aspect of our training.

Conclusion

It is likely that the mere passage of time—and the continued social advances one can expect to accompany it—will resolve many of the current issues related to women in combat. As Americans' views toward traditional male and female roles continue to change, our society's views toward women in combat will probably change as well. In the meantime, however, behavioral scientists should conduct detailed studies regarding the likely consequences of women serving in every aspect of combat—including the prisoner of war environment. Through such studies of women as prisoners of war, behavioral scientists would provide an invaluable service to our nation's policymakers as they consider this one very important aspect of women in combat.

References

Burke, K. & Lin, T. (1988, August). A Systematic Laboratory Approach to Amenorrhea. Physician Assistant, 106.

Guzman, A. (1989, November 21). Personal interview with Nurse Practitioner, Cadet Clinic Primary Care, United States Air Force Academy.

Predictors of Equal Opportunity Training Performance and Citizenship Behaviors[1]

L. Alan Witt, Ph.D.
Western Illinois University
LtCol Mickey R. Dansby, Ph.D.
Defense Equal Opportunity Management Institute

Abstract

The present study examined person- and situation-level predictors of equal opportunity citizenship behaviors (EOCBs) and training performance grade-point average of 96 students in the 16-week resident equal opportunity course of the Defense Equal Opportunity Management Institute, Patrick Air Force Base, Florida. Data indicated that personal views about equal opportunity and dispositional positive affect accounted for variance in EOCBs, while equity perceptions accounted for variance in training performance.

Organizations rely on their personnel to help co-workers with problems, tolerate inconveniences without complaint, promote a positive work climate, and protect organizational resources. These "prosocial organizational behaviors" (Brief & Motowidlo, 1986) or "organizational citizenship behaviors" (OCBs; Organ, 1986) are discretionary and aimed at improving the welfare of co-workers, the work group, or the organization. While they are neither prescribed nor required in advance for a job, they may have a significant impact on organizational effectiveness.

Research on OCBs has focused on behaviors of traditional interest to private sector managers and ignored socially responsible behaviors of considerable interest to the Department of Defense (DoD), namely behaviors that enhance equal opportunity (EO) and equitable treatment of individuals. At the Defense Equal Opportunity Management Institute (DEOMI), military personnel are trained to serve as EO advisors, and part of that training involves EO citizenship behaviors (EOCBs). Given the importance of this type of socially responsible behavior, a theoretical and practical issue becomes the identification of factors that give rise to or inhibit these behaviors in the military setting. In other words, what factors motivate a person to manifest EOCBs? In line with Brief and Motowidlo's (1986) call for the identification of the individual-level correlates of prosocial organizational behavior and with the DoD's concern with EO behaviors, the present study examines both situation- and person-level predictors of EOCBs and EO training performance. Variables include perceptions of the EO climate, perceptions of equity, situational affect (i.e., job satisfaction), dispositional affect, and personal views about EO.

Both practitioners and researchers have recognized the impact of organizational climate on EO and affirmative action (e.g., Sargent, 1978). In line with Scheinfeld and Zalkind's (1987) work demonstrating the impact of

[1]This paper reflects the opinions of the authors only and should not be construed as official policy of the Department of Defense or any organization.

civil liberties climate on organizational outcomes, Landis and Fisher (1987) argued that individual perceptions of an organization's "equal opportunity climate" have a direct impact on mission effectiveness in military organizations. They (1987, p. 8) defined EO climate as the expectation of organization members "that they will have equal access to opportunities, responsibilities, and rewards within an organization." Individuals perceiving greater EO at their duty stations (i.e., having a more favorable psychological climate for EO) should perceive EO as a norm. Thus, we predicted they would engage in greater levels of EOCBs and work harder to do well in EO training.

Procedural justice research on organizational rewards has provided evidence of the importance of resource distribution procedures as determinants of fairness in organizations. For example, Landy, Barnes-Farrell, and Cleveland (1980) found that the process by which workers' performance appraisals were determined was related to the perceived fairness of their evaluation, independent of outcome. Similarly, Greenberg (1987) reported that laboratory-manipulated fairness of procedures influenced perceptions of the fairness of outcomes. In line with the literature suggesting an exchange basis for responses to equity/inequity, we predicted that individuals perceiving equity (i.e., those who perceive they have been treated fairly) with regard to outcomes such as duty assignments, fitness reports, and overall treatment would engage in greater levels of EOCBs and work harder in EO class.

Considerable evidence suggests that mood is a strong predictor of prosocial behavior (Rushton, 1984). Organ (1988, p. 34) argued, "If we . . . assume that job satisfaction translates into a predominant mood state . . . , then it follows . . . that job satisfaction scores should bear a positive correlation to a valid index of OCB." Congruent with the assumption that job satisfaction reflects dispositional affect, researchers have used job satisfaction as a surrogate for dispositional affect in studies examining the relationship between affect and OCBs. Organ has suggested that individuals seek to reciprocate the organization to the extent that the organization benefits them (i.e., to the extent that they are satisfied with the organization). Therefore, the more satisfied the individual is, the more likely the individual will "pay back" the organization with OCBs. Organ has noted that, while an individual may not have the ability to provide innovative solutions to work problems, he or she may have the capacity to manifest good citizenship behaviors. He (1988, p. 43) cited eight studies in support of his argument and suggested that, "they lend reasonable support to the hypothesis that OCB and job satisfaction are bound together." The emerging OCB literature suggests that job satisfaction may have a direct bearing on the extra-role aspects of job performance. Following this literature, we predicted that personnel more satisfied with their jobs would be more likely to engage in greater levels of EOCBs and work harder in the EO course.

Staw, Bell, and Clausen (1986, p. 61) argued that employees "bring a positive or negative disposition to the work setting, process information about the job in a way that is consistent with that disposition, and then experience job satisfaction or dissatisfaction as a result." They asserted (p. 61) that the development of job attitudes "come[s] as much from the internal state of the individual as from any external cues." In line with their argument, is it context-relevant job satisfaction or dispositional affect that is related to OCBs? In the studies cited by Organ (1988), affect was operationalized as job satisfaction, although research has shown that job-related stressors accounted for about 4% of the variance in OCBs.

In studies of self-reported mood, negative and positive affect consistently have been identified as the two relatively independent and dominant dimensions (Watson & Tellegen, 1985). Watson, Clark, and Tellegen (1988, p. 1063) defined them:

> Positive Affect (PA) reflects the extent to which a person feels enthusiastic, active, and alert. High PA is a state of high energy, full concentration, and pleasurable engagement, whereas low PA is characterized by sadness and lethargy. In contrast, Negative Affect (NA) is a general dimension of subjective distress and unpleasurable engagement that subsumes a variety of aversive mood states, including anger, contempt, disgust, guilt, fear, and nervousness, with low NA being a state of calmness and serenity.

This independence means that High and Low PA can have different patterns of relationships with other variables. As noted by Brief and Motowidlo (1986), this helps to explain some of the inconsistent findings in the altruism literature. While data suggest that positive moods generally promote helping behavior (Rushton, 1984), negative moods have been found to both increase and decrease it (Clark & Isen, 1982). Recent advances in affect theory and methodology permit assessment of both positive and negative affect in organizational settings. In line with this body of research, we predicted that individuals experiencing greater positive affect and job satisfaction and lower negative affect would engage in greater levels of EOCBs and work harder to learn means to communicate EO concerns to others.

We were also interested in individuals' personal views about EO. We predicted that individuals who held views more supportive of EO in the military would be more likely to engage in greater levels of EOCBs and work harder to learn means to communicate EO concerns to others.

Method

Ninety-six volunteers from the 122 students in Class 89-3 of the resident EO training course at DEOMI anonymously completed a questionnaire in response to a request. Predominantly, participants were male U.S. Army enlisted members in the grade of sergeant first class (E-7). For the most part, the students were not volunteers for EO duty. Equal Opportunity Citizenship Behaviors were assessed by seven items ($\bar{X} = 27.4$, $SD = 4.5$, alpha = .66). Training performance grade-point average was assessed by one question, in which students were asked to indicate their overall grade-point average. Psychological climate for EO was measured by the 13 items loading highest on the Landis and Fisher (1987) EO climate measure ($\bar{X} = 44.6$, $SD = 13.2$, alpha = .93). Perceptions of fairness of fitness reports, duty assignments, job tasks assignment, and overall treatment in the military (i.e., equity perceptions) were assessed by 12 items ($\bar{X} = 61.8$, $SD = 8.5$, alpha = .91). Eight items measured job satisfaction ($\bar{X} = 32.1$, $SD = 6.8$, alpha = .85) Positive affect and negative affect were measured by the Watson et al. (1988) 10-item Positive Affect Scale ($\bar{X} = 38.6$, $SD = 4.6$, alpha = .85) and 10-item Negative Affect Scale ($\bar{X} = 20.5$, $SD = 6.8$, alpha = .94), respectively. Personal beliefs about EO were measured by one item in which students were asked to indicate the extent to which they held beliefs supporting EO efforts in the military ($\bar{X}$ = 4.3, SD = .88).

Results

The correlations between EOCBs and the variables of interest were as follows: perceptions of EO climate, r = -.10, ns; perceptions of equity, r = -.15, ns; job satisfaction, r = .20, p < .03; dispositional positive affect, r = .39, p < .01; dispositional negative affect, r = -.22, p < .02; personal beliefs about EO, r = .41, p < .01; and training performance grade-point average, r = -.10, ns.

The correlations between training performance grade-point average and the variables of interest were as follows: perceptions of EO climate, r = -.11, ns; perceptions of equity, r = -.42, p < .01; job satisfaction, r = .02, ns; dispositional positive affect, r = .11, ns; dispositional negative affect, r = -.06, ns; and personal beliefs about EO, r = -.05, ns.

Discussion

Before discussing the data, we emphasize limitations in the present study. First, the sample was one of convenience; second, the data may be subject to method variance; third, because there was no way the individuals could be identified, we were unable to gather archival performance data.

The finding that positive affect accounted for a small amount of the variance in EOCBs suggests that individuals dispositionally high is positive affect are more likely to engage is EOCBs. Moreover, it is consistent with the mood-altruism literature. The failure of job satisfaction to account for variance in EOCBs questions the utility of situational affect as an antecedent of citizenship behaviors and is in contrast to work by Organ and his colleagues (cf. Organ, 1988).

The small amount of variance in EOCBs explained by views about EO suggests that individuals who believe in EO are more likely to act consistent with their beliefs. This reinforces the need for EO training to focus on attitude change among individuals holding views not supporting EO.

The finding that equity perceptions were negatively related to grade-point average was unexpected. These data may suggest that individuals who perceive they have been treated unfairly may work harder in EO training. Of course, a number of other person-level variables, such as cognitive complexity, writing ability, and previous training in the topic areas should be examined to determine whether it is equity or some other variable that is accounting for this variance in training performance.

Findings from this study may have both theoretical and practical implications. On a theoretical level, the findings may provide a working framework to explain differences in persons' EOCBs. On a practical level, EOCBs may provide an alternative means of validating work by Landis and Fisher (1987). Because the individuals came from different duty stations, the climate for EO measure may have been less salient in the development of the decision to manifest EOCBs among individuals in this sample. However, the measurement of psychological climate for EO described by Landis and Fisher may provide a diagnostic method of assessing a group or organization's norms for EO outcomes. The empirical norms may then be compared with desired ones to pinpoint areas in which organizational development efforts are needed. The measure of climate for EO may also be used to make inferences regarding change subsequent to intervention.

References

Brief, A.P., & Motowidlo, S. J. (1986). Prosocial organizational behaviors. Academy of Management Review, 11, 710-725.

Clark, M. S. & Isen, A. M. (1982). Toward understanding the relationship between feeling states and social behavior. In A. H. Hastorf & A. M. Isen (Eds.), Cognitive social psychology (pp. 73-108). New York: Elsevier-North Holland.

Greenberg, J. (1987). Reactions to procedural injustice in payment distributions: Do the means justify the ends? Journal of Applied Psychology, 72, 55-61.

Landis, D., & Fisher, G. (1987). Construction and preliminary validation of an instrument to measure equal opportunity climate. Contract No. F49620-85-C-0013. Air Force Office of Scientific Research.

Landy, F. J., Barnes-Farrell, J., & Cleveland, J. N. (1980). Perceived fairness and accuracy of performance evaluation: A follow-up. Journal of Applied Psychology, 65, 355-356.

Organ, D. W. (1988). Organizational citizenship behaviors: The good soldier syndrome. Lexington, MA: Lexington.

Rushton, J. P. (1984). The altruistic personality: Evidence from laboratory, naturalistic, and self-report perspectives. In E. Staub, D. Bar-Tal, J. Karylowski, & J. Reykowski (Eds.), Development and maintenance of prosocial behavior: International perspectives on positive morality (pp. 271-290). New York: Plenum.

Sargent, A. G. (1978). Affirmative action: Total systems change. University of Michigan Business Review, 30, 18-26.

Scheinfeld, D., & Zalkind, S. S. (1987). Does civil liberties climate in organizations correlate with job satisfaction and work alienation? Psychological Reports, 60, 467-477.

Staw, B. R., Bell, N. E., & Clausen, J. A. (1986). The dispositional approach to job attitudes: A lifetime longitudinal test. Administrative Sciences Quarterly, 31, 56-77.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. Psychological Bulletin, 98, 219-235.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. Journal of Personality and Social Psychology, 54, 1063-1070.

Initial Validation of a Personnel Selection System For Landing Craft Air Cushion (LCAC) Vehicle Operators

Daniel L. Dolgin, Ph.D.

and

Tatree Nontasak, Ph.D.

Naval Aerospace Medical Research Laboratory
Pensacola, Florida 32508-5700

Abstract

To support the performance-based selection of LCAC operators, we identified and evaluated a battery of personality, cognitive, and psychomotor tests at the Naval Aerospace Medical Research Laboratory. Initial results indicate that several tests are predictive of LCAC training criteria. Twenty-six LCAC trainees took the test battery prior to their entry into the LCAC training program. We collected and analyzed training performance data from the trainees at the end of the course of training. Descriptive statistics and correlation coefficients were derived. Psychomotor and risk test performance demonstrate a significant relationship with the LCAC performance criteria. Initial results indicate that the use of computer-based testing will improve the selection of LCAC operators. The implications of these findings for LCAC personnel selection are discussed.

Introduction

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

As training costs escalate and projected plans call for an increased number of LCAC vehicles and crews, selecting candidates who will be successful in the operation of complex machinery becomes more critical. In 1988, the Naval Aerospace Medical Research Laboratory (NAMRL) completed the development of a computer-based performance assessment battery for LCAC. This automated battery assesses basic information processing abilities, higher order processes, psychomotor skills, time-sharing ability, and personality traits that might predict success in LCAC training. The test battery was based on previous literature and research from pilot selection, e.g., Bordelon and Kantor (1986); Damos and Gibb (1986); Dolgin and Gibb (1988).

A concurrent validity study examining Navy fleet LCAC operator performance on the automated test battery was reported recently (Nontasak, et al. 1989). In that study, several predictor tests were significantly correlated with measures of success in LCAC training. This report contains the results of an initial predictive validation of the test battery for LCAC trainees. Descriptive statistics and individual correlations between test measures and LCAC training criteria are presented. In addition, the relationship between biographical information and LCAC training criteria was evaluated.

Two criterion measures were used: underway grade (UG) and primary training grade (PTG). The UG is the core criterion and represents a composite score reflecting a student's tactical performance in the training hovercraft. The PTG measure consists of 40% classroom grade with the remainder reflecting UG scores. We did not relate test battery measures to specific components of LCAC training. Past research in flight training has shown the difficulty of identifying reliable subcriteria embedded within the more global criteria (Dolgin, et al. 1987a).

Methods

Subjects

Student LCAC trainees (N = 26) were selected for LCAC training on the basis of current medical examinations and preference. Subjects participated voluntarily and were informed that the investigation involved performing tasks in problem solving and perceptual and motor skills. Subjects were briefed that their test performance would not affect their continuation in the program or be entered into their permanent service records. Results would be used solely to develop an improved selection system for the Navy. The subjects were 20 to 42 years old (M = 31.19, SD = 5.93). All candidates were male and had successfully completed a minimum of a high school education. Classification variables were recorded for each subject and included: 1) age, 2) military rating/classification, 3) amount of time in the military, and 4) geographical accession area.

Apparatus

All testing was conducted on Apple IIe microcomputers with control sticks and foot pedals. Subjects used an Apple IIe numeric keypad to respond to discrete stimuli. All responses were recorded to millisecond accuracy. For psychomotor tests (PMT), two control sticks (Measurement Systems, Inc., 542) were used for cursor control during the tracking tasks. One control stick was mounted in the center on the forward edge of a standard straight-back metal chair. The other stick (throttle) was mounted on the left edge of the chair. Additional apparatus included rudder pedals patterned after those of a Systems Research Laboratory psychomotor test device. Two Jameco JE 520-AP Voice Synthesizers were used to present the Dichotic Listening (DLT) letter-digit strings over binaural headphones. Further details of the test battery are included in Dolgin, et al. (1987b) for the Risk test; Damos and Gibb (1986) for the One-dimensional Compensatory Tracking and Manikin tests; and Griffin (1987) for the PMT and DLT tests.

Procedure

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

Results

Psychomotor Task

Performance on the series of psychomotor multitask conditions correlated with training criteria. Table 1 lists the tests that correlated significantly with training criteria. When performed in combination with the DLT, PMT tasks indicated a moderate relationship to both UG and PTG with coefficients ranging from -.41 to -.48. As expected, the UG correlations were higher than those derived for PTG because UG reflects the eye-hand-foot coordination segment of PTG. These correlations were in the expected direction. That is, greater psychomotor tracking error was associated with lower overall primary training grades.

The error scores from the combined stick-throttle-rudder tasks, without the DLT, only yielded a correlation of -.20 with PTG. When UG was used as the criterion, the correlation was slightly higher (-.22). Though not statistically significant, these were also in the expected direction.

Risk Test

When personality attributes were analyzed, the risk test was significantly related to PTG but not to UG. The number of boxes selected was significantly related to the PTG criterion ($r = .46$, $p < .05$, $n = 26$). Response time on the test, however, did not demonstrate a significant relationship with either training criterion.

Age

Subjects' age indicated significantly negative correlations to both UG and PTG ($r = -.40$ and $-.46$, $p < .05$, $n = 26$). The older the subject, the greater the decrement in LCAC training performance.

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

Test measure	Underway grades	Primary training grades
PMT (error score)		
Stick-rudder + DLT	-.48*	-.44*
Stick-throttle + DLT	-.46*	-.41*
Stick-rudder-throttle + DLT	-.46*	-.40
Risk taking		
No. of responses	.36	.46*
Age	.40*	-.46*

* $p < .05$, two-tailed

One-dimensional Compensatory Tracking

The average cumulative distance off target (CDOT) error for the final three trials of the task was used because it is stable. Average CDOT error was $M = 27.97$, $SD = 13.74$ for LCAC trainees ($n = 26$). Both UG and PTG correlated equally with average CDOT ($r = -.20$). When the one-dimensional compensatory tracking task was performed in combination with the digit cancelation task, the correlation between average CDOT and UG remained at $r = -.20$. The correlation between average CDOT and PTG was $r = -.25$. None of these correlation coefficients was significant at the .05 level.

Dichotic Listening Task

None of the four DLT measures indicated significant correlation to the training criteria. Single-DLT demonstrated a weak correlation to UG ($r = .17$) and to PTG ($r = .13$). The three multi-DLT measures yielded no relationship to either criterion.

Manikin Test

The correct reaction times for the last four trials of the Manikin task were stable and were averaged for use as the independent measure. Mean (SD) correct reaction time was 1.74 with an SD of .53 ($n = 24$). Mean correct reaction time on the task correlated $r = .17$ with both UG and PTG ($n = 24$). Manikin number correct ($M = 79.04$, $SD = 19.45$, $n = 24$) correlated ($r = .14$, $n = 24$) with UG and PTG ($r = .12$, $n = 24$). The Manikin test did not display any significant relationship to the training criteria.

Discussion

The psychomotor task performed in combination with the dichotic listening task were significantly related to both the UG and PTG criteria. Of the two risk test measures (number of boxes selected and response time), the number of boxes selected correlated significantly with the LCAC training criterion. Our results also showed that age was significantly related to the successful completion of the LCAC training program. These results provide support for the notion that certain abilities and personality traits (as measured by several tests) may be useful in screening individuals for the LCAC training program.

Subjects who perform better (lower error score) on the multitask tests appear to have a greater likelihood of success in the training syllabus. Those subjects with a moderate risk-taking propensity also display the characteristics that appear to facilitate completion of training. Additionally, younger subjects tend to perform better in the program. Taken together, these characteristics should be considered important in the initial LCAC personnel screening process.

The variables identified in our study appear promising for more accurately predicting 1) those who will fail from training, and 2) actual LCAC performance in the primary portion of tactical, underway training. The latter is very important in that the ability to predict who will perform well in LCAC training enables pre-identification of the successful trainees.

Due to our relatively small sample size, these results should be considered preliminary with minimal generalization. The current findings represent the promise of an improved LCAC selection system that may both decrease accidents and increase the quality of LCAC operators in the U.S. Navy. Continued research and evaluation of the valid tests in the battery with an LCAC-trainee population should provide considerable advances over present screening techniques.

Acknowledgments

This work was sponsored by NAVSEA (Code PMS 377) through Naval Ocean Systems Center and supported by Naval Coastal Systems Center (Work Unit Number CE02). The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of the Navy, Department of Defense, nor the U.S. Government. Volunteer subjects were recruited, evaluated, and employed in accordance with the procedures specified in the Department of Defense Directive 3216.2 and Secretary of the Navy Instruction 3900.39 series. These instructions are based upon voluntary informed consent and meet or exceed the provisions of prevailing national and international guidelines. Trade names of materials and/or products of commercial or nongovernment organizations are cited as needed for precision. These citations do not constitute official endorsement or approval of the use of such commercial materials and/or products. The authors wish to acknowledge the assistance of Peter Collyer for his programming effort and Solomon Eagles for test administration.

References

Bordelon, P.V. and Kantor, E.J. (1986). Utilization of psychomotor screening for USAF pilot candidates: Independent and integrated selection methodologies (AFHRL-TR-86-4). San Antonio, TX: Air Force Human Resources Laboratory.

Damos, D.L. and Gibb, G. D. (1987). Development of a computer-based naval aviation selection test battery (NAMRL-1319). Pensacola, FL: Naval Aerospace Medical Research Laboratory. (DTIC No. AD A179 997)

Dolgin, D.L., Gibb, G.D., Nontasak, T., and Helm., W.R. (1987a). Instructor pilot evaluations of key naval primary flight training criteria (NAMRL-133[illegible]). Pensacola, FL: Naval Aerospace Medical Research Laboratory. (DTIC NO. AD A183 793)

Dolgin, D.L., Shull, R.N., and Gibb, G.D. (1987b). Risk assessment and the prediction of student pilot performance. In R.S. Jensen (Ed.), Proceedings of the Fourth International Symposium on Aviation Psychology, Columbus, OH, April, 480-485.

Dolgin, D.L. and Gibb, G.D. (1988). A review of personality assessment in aircrew selection (NAMRL Monograph 36). Pensacola, FL: Naval Aerospace Medical Research Laboratory. (DTIC No. AD A200 392)

Griffin, G.R. (1987). Development and evaluation of an automated series of single and multiple dichotic listening and psychomotor tasks (NAMRL-1336). Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Nontasak, T., Dolgin, D.L., and Griffin, G.R. (1989). Performance-based tests, personality attributes, and training outcome among landing craft air cushion (LCAC) vehicle operators. Proceedings of the Human Factors Society 33rd Annual Meeting, Denver, Colorado, October, 901-904.

MODELING CAREER PATHS IN AIR FORCE ENLISTED SPECIALTIES

James R. Knight, David S. Vaughan,
& Robert M. Yadrick
McDonnell Douglas Missile Systems Company
St. Louis, MO & San Antonio, TX

Winston R. Bennett
Training Systems Division
Air Force Human Resources Laboratory
Brooks Air Force Base, TX

ABSTRACT

Present military occupational analysis methodologies focus on the identification of job types within an occupation and analysis and evaluation of initial skills training programs. New methods have recently been developed which help identify career paths including job assignment probabilities and associated training required for job mastery. By integrating Job and Training History survey information with job analysis data, a more comprehensive model of occupational career paths can be developed. Such dynamic models of specialties represent a significant enhancement of the current job analysis technology.

Recent developments in occupational analysis and training research, as well as in Air Force decision making processes, have created new opportunities for optimizing manpower, personnel, and training (MPT) planning and decision making. Such developments include the recent emergence of Utilization and Training Workshops (U&TWs) and Training Planning Teams (TPTs) as the primary vehicles for making major training decisions. Such innovative procedural changes also make obvious a need for a technologically advanced data generation, analysis, and evaluation capability. To make good MPT decisions about an AFS or weapons system, decision makers must be able to visualize and understand the jobs and training programs of the AFS or weapon system under consideration, as well as the relative costs and payoffs of various training options. Such a "model" of an AFS provides a concise summary of its current status, creates a common "language" for discussion or negotiation, and forms the baseline against which various proposals or options can be evaluated.

The Air Force presently uses a task-based approach to evaluate initial skills technical training content and review personnel classification and utilization policies (Christal, 1974; Mitchell, 1988). As part of the occupational analysis (OA) process, tasks are defined by subject-matter experts (SMEs) of an AFS in their own technical terminology, working with analysts of the USAF Occupational Measurement Center, Randolph Air Force Base, Texas (see AFR 35-2). Several kinds of data on these tasks are collected from job incumbents and supervisors for use in reviewing training programs (see ATCR 52-22). Large samples of incumbents are asked to provide information about which tasks they perform in their present jobs and the relative amount of their job time spent performing such tasks. These data are used to examine the variety of specialized jobs within an AFS (occupation), to assess how jobs change at advanced skill levels, and to review official AFS descriptions and initial training programs (Christal & Weissmuller, 1988; Mitchell, Ruck, & Driskill, 1988).

TECHNICAL APPROACH

To provide adequate support for advanced training decision making, the Air Force Deputy Chief of Staff for Personnel, Education and Training (HQ USAF/DPPE) requested that the Air Force Human Resources Laboratory (AFHRL) develop a computer-based Training Decisions System (TDS) to augment the Air Force ISD model. This prototype system generates necessary front-end training requirements data, validated decision algorithms, and procedures for improved interaction among training, personnel, and functional managers. The TDS focuses on supporting Air Force managers in making decisions as to the what, where, and when of the technical training (including the OJT) required for an AFS (Ruck, 1982; Vaughan, Mitchell, Yadrick, Perrin, Knight, Eschenbrenner, Rueter, & Feldsott, 1989).

The TDS meets the requirement for a technologically advanced MPT decision support system. The TDS includes procedures for building data bases concerning the dynamic flow of people through jobs and concerning both the formal training and OJT required to support a dynamic job flow. Furthermore, the TDS includes modeling and optimization capabilities which provide estimates of training quantities, costs, and capacities for both formal training and OJT. These modeling capabilities allow the TDS to go beyond simply describing the current situation in an AFS--TDS can model alternative scenarios which reflect possible MPT policy options. This modeling capability is used to study trade-offs between formal training and OJT for meeting particular task training requirements.

For TDS, a model of the specialty being studied is developed in the form of a graphic flow diagram of the current Utilization & Training (U&T) pattern (see Figure 1 for an example), as well as narrative summaries (for descriptive purposes and later use in surveying the U&T pattern preferences of managers; see Yadrick, Knight, Mitchell, Vaughan, & Perrin, 1988). Such a graphic display provides some sense of the flow of individuals through training programs and jobs but does not lend itself to summarizing the various types of quantitative information involved, such as the number of individuals entering the field each year, or the various probabilities of reassignment among jobs, attending advanced technical training, or participating in Professional Military Education (PME) courses.

One of the things needed to create a dynamic picture of a specialty, is data on the assignment probability to each job and to each training program. Such data are needed to understand the flow of personnel from one job to another which generates new OJT requirements each year, as well as calculate attendance at FTDs, PME courses, and other training programs. Personnel records have some of this information but job assignments are not consistently recorded in a way which can be translated to recognized OSR job types (except where a job title includes such designators as "NCOIC").

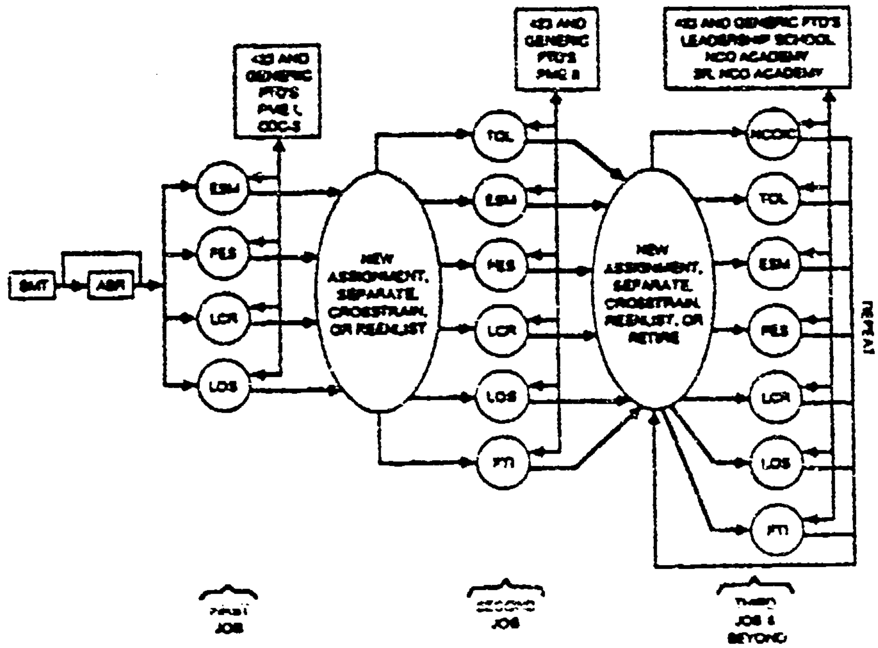


Figure 1. Current U&T Pattern; Environmental Systems Maintenance (AFS 423X1)

For first job personnel (defined generally as individuals in their first 24 months in the specialty), the probability of being assigned to a given job is equal to the proportion of people with 1 - 24 months Total Active Federal Military Service (TAFMS) in those jobs, which we can approximate from the OSR data. This is complicated somewhat by attrition out of the specialty (or out of the Air Force) and crossflow from other specialties. What we really need is a more longitudinal picture of how people move from job to job, or from given jobs out of the career field.

To develop this kind of information, we designed a Job and Training History survey (JTHS) which asked present job incumbents to indicate which jobs they had held at what points in time, as well as which courses they had attended. The sample was selected from the OSR data files, to insure that we had all of the job groups identified in the OSR represented in the sample. A separate random sample of new recruits (1 - 48 months TAFMS) was developed, to assess any recent trends in assignment policy. The survey was administered through local base Survey Control Officers.

For the Environmental Systems Maintenance specialty, AFS 423X1, the JTHS form included a background section and a listing of nine jobs with space to indicate dates of assignment; each job was defined in terms of its characteristic Task Modules. Separate sections of the survey asked for dates of attendance in any formal training program. Job incumbents were encouraged to fill in any courses not listed. The data collected appeared realistic and proved very useful; a brief validity check against known school attendance rates indicated good correspondence between data sets.

Once the average length of assignment is determined from the JTHS, the ASCII CODAP Group Membership Selection (MEMSEL) program is used to sort cases into job type membership per average assignment period. Other CODAP routines are used to calculate the proportion of total cases in the assignment period per cell (each job type per average assignment period). The product is an ordered array of numbers interpretable as assignment probabilities at each career point, as represented by TAFMS periods.

With this kind of distributional data, it is possible to flesh out the graphic display (shown as Figure 1) of how jobs and training programs relate to one another throughout a career in the specialty. The graphic display is an effective communicative tool, but does not lend itself to summarizing the quantitative information involved in modeling a U&T pattern, such as the number of individuals entering the field each year, or the probabilities of reassignment among jobs, attending advanced technical training, or leaving the AFS.

In the TDS, these kinds of data are used in a dynamic simulation program to estimate annual flows of individuals through jobs and training activities, including the hours of on-the-job training (OJT) required for each person to master their particular job. Such a simulation permits us to estimate the costs of OJT required in a year as well as formal training costs. Likewise, such a simulation permits us to estimate the training capacity of representative field units (see Vaughan, et al., 1989).

This very flexible, Task Module-based approach to estimating training requirements and job assignment probabilities within an Air Force specialty also facilitates the examination of possible changes to AFS jobs or associated training programs. The TDS deals with such proposed changes as an Alternative U&T pattern. Once the details of how such a change would impact job content and training programs, the simulation program can be rerun so as to estimate training requirements under the alternative configuration. These are used to calculate training cost and capacity estimates for the alternative, which are compared to present costs to evaluate the possible impacts of the proposed changes.

Figure 2 illustrates an example of an alternative configuration for AFS 423X1 which was suggested by some senior NCOs as a potentially valuable change. In this alternative, jobs and training programs are specialized by large and small aircraft, so individuals can develop a full career path on one type of system or the other (but not be required to maintain both).

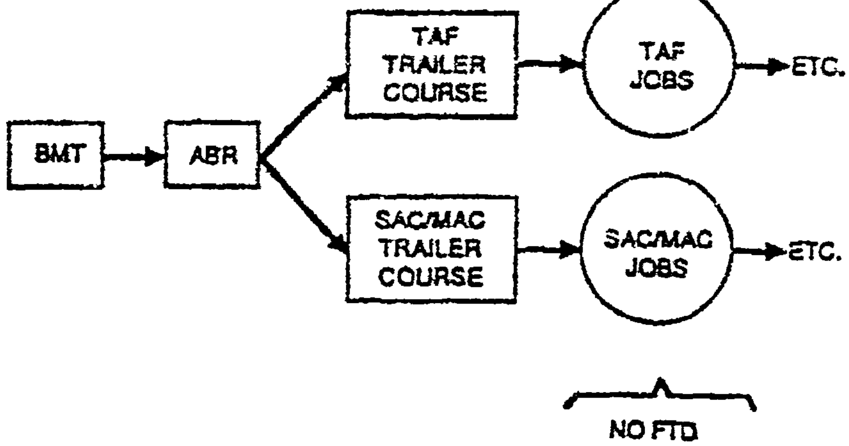


Figure 2. Alternative U&T Pattern for Specializing AFS 423X1 Jobs.

Cost estimates for this alternative include $2.4 million for formal courses and $550 thousand for OJT (versus $1.9 million for courses and $530 thousand for OJT in the current U&T pattern). Such specialization would be more expensive (roughly half a million dollars per year) but does not exceed the OJT capacity of field units. Thus, it is a feasible option, but only if additional funding can be obtained.

DISCUSSION

Present occupational analysis methodologies focus primarily on the identification of job types and assessment of initial skills training programs. New TDS methodologies, such as Job and Training History surveys and the U&T simulation program, now make it possible to examine the dynamics of a specialty, identify career paths, estimate job assignment probabilities, and calculate the hours of training required for job mastery. By integrating JTHS information with OA data, a more comprehensive model of an AFS can be developed. Further, it is now possible to assess the impact of proposed changes in an AFS in terms of total training costs and training capacity constraints. Such models represent a significant enhancement of the current job analysis technology, and have the potential for substantially improving MPT planning and decision making.

REFERENCES

Air Force Regulation 35-2 (1982, 23 July). Occupational analysis. Washington, DC: Headquarters, United States Air Force.

Air Training Command Regulation 52-22 (1981, 16 October). Occupational analysis program. Randolph AFB, TX: HQ, Air Training Command.

Christal, R.E. (1974). The United States Air Force occupational research project (AFHRL-TR-73-75, AD-774 574). Lackland AFB, TX: Occupational Research Division, Air Force Human Resources Laboratory.

Christal, R.E., & Weissmuller, J.J. (1988). Job-task inventory analysis. In S. Gael (Ed), Job analysis handbook for business, industry, and government. New York: John Wiley and Sons, Inc. (Chapter 9.3).

Mitchell, J.L. (1988). History of job analysis in military organizations. In S. Gael (Ed), Job analysis handbook for business, industry, and government. New York: John Wiley and Sons, Inc. (Chapter 1.3).

Mitchell, J.L., Ruck, H.W., & Driskill, W.E. (1988). Task-based training program development. In S. Gael (Ed), Job analysis handbook for business, industry, and government. New York: John Wiley and Sons, Inc. (Chapter 3.2).

Ruck, H.W. (1982, February). Research and development of a training decisions system. Proceedings of the Society for Applied Learning Technology. Orlando, FL.

Vaughan, D.S., Mitchell, J.L., Yadrick, R.M., Perrin, B.M., Knight, J.R., Eschenbrenner, A.J., Rueter, F.H., & Feldsott, S. (1989, June). Research and development of the training decisions system (Final Report; AFHRL-TR-88-50). Brooks AFB, TX: Prepared for the Training Systems Division, Air Force Human Resources Laboratory.

Yadrick, R.M., Knight, J.R., Mitchell, J.L., Vaughan, D.S., & Perrin, B.M. (1988, July). Training decisions system: Development of the field utilization subsystem (AFHRL-TR-88-7). Brooks AFB, TX: Training Systems Division, Air Force Human Resources Laboratory.

USAF Academy Graduates in Logistics

Captain William W. Roberts
Air Force Military Personnel Center

Abstract

At the 1988 base-level logistics conference held at the United States Air Force Academy (USAFA), senior logisticians discussed the need to improve methods for informing officer candidates, particularly at the Academy, of the challenges and rewards of a career in logistics. Based on Air Force force characteristics, USAFA graduates are underrepresented in the logistics career field. The Air Force Logistics Management Center (AFLMC) and the Air Force Military Personnel Center (AFMPC) designed and administered a survey to USAFA graduates serving in the logistics career field (N=472). The survey's purpose was to determine perceptions of logistics and associated USAFA curriculum junior and senior officers had as cadets. Additionally, we identified current satisfaction of USAFA graduates regarding the logistics career field. While attitudes regarding the career field were somewhat negative for the USAFA cadet, there is satisfaction with a career in logistics after working in the career field.

At the 1988 base-level logistics conference held at the United States Air Force Academy (USAFA), senior logistics function representatives (primarily in the grade of Colonel) discussed USAFA graduate representation in logistics and how the Academy prepares graduates for a logistics career. Some believed that Academy graduates were not represented adequately in logistics career fields and that the graduates were not adequately prepared by the Academy to enter the logistics career field. Others felt Academy cadets did not perceive a logistics assignment as favorable. At the conclusion of the conference, the Air Force Logistics Management Center (AFLMC) was tasked to investigate these concerns and submit a report on USAFA graduates in logistics.

Method

Subjects

Surveys were sent to 472 Academy graduates currently serving in a logistics assignment (004X, 008X, 009X, 31XX, 40XX, 60XX, 62XX, 64XX, 65XX, 66XX) as of March, 1989. Of 472 officers, 362 completed and returned the survey for a response rate of 77%. Forty-nine percent are field-grade and 51% are company-grade officers. Thirty-five percent entered logistics on a direct duty assignment from the Academy, 22% are flight training eliminees, and 32% entered logistics later via career broadening, cross training, or as a rated supplement assignment.

Apparatus

The survey was organized into three sections; Background, Academy Preparation, and Job Satisfaction. The Background section enhances analysis of the overall survey based on factors such as grade, gender, and major command. Academy Preparation centers on the perceptions graduates had as cadets of how the Academy prepared them for logistics careers such as the amount, sufficiency, and source of logistics information presented. Respondents answered Job Satisfaction questions from a present perspective of their logistics jobs and addressed issues such as promotion potential, working conditions and desire to transfer to another AFSC.

Procedure

Force characteristics were obtained from the Air Force Military Personnel Center (AFMPC) to identify the representation of USAFA, OTS and ROTC graduates throughout the Air Force and within logistics as of March, 1989. Policy documentation was researched to determine some of the reasons for current force structure characteristics as indicated by demographic data. Also, in order to obtain an assessment of perspectives, the AFLMC and AFMPC developed and administered an attitude and opinion survey to all USAFA graduates in logistics career fields and analyzed the returned data.

Results

Force characteristic data indicate that 13% of Air Force officers have come from the Academy. Yet, only five percent of individuals in the logistics career field are Academy graduates. There are several reasons for this underrepresentation. By comparing percentages of Academy graduates in each career area to the percentage of graduates in the entire Air Force (13%), we can get a better idea of just how well Academy graduates are represented in logistics. The results of this reveal that International Political and Military Affairs (19%), Operations (18%), and Scientific and Developmental Engineering (16%) have a proportion greater than 13%, while logistics career areas have only 5% Academy graduate representation (Roberts, 1989). Primary reasons for this are the requirement for at least 65% of each graduating class to enter Undergraduate Pilot Training (UPT) and lack of logistics education/emphasis. While these are causes for present proportions, recent Air Force policies and procedures (Boles, 1989; Hamm, 1989; Hickey, 1989) will most likely decrease Academy representation even further in the future.

Air Force policy states that the majority of Academy cadets (at least 65%), immediately upon graduating, enter Undergraduate Pilot Training (UPT) and remaining graduates receive Category A, B1, B2, or C functional-area assignments. Two logistical career fields, 31XX (Missile Maintenance) and 40XX (Aircraft/Munitions Maintenance) are Category A assignments, and five logistical career fields, 60XX (Transportation), 62XX (Services), 64XX (Supply Management), 65XX (Acquisition Contracting/Manufacturing) and 66XX (Logistics Plans and Programs) are Category B1 assignments. Most Academy graduates who do not immediately enter UPT receive A, B1, or B2 assignments. The assignment process has recently changed for graduates initially entering UPT but not finishing. Under the previous policy, these officers would be reassigned to a Category A, B1, B2, or C position. The new policy is to reassign them to only Category A positions. This means that 31XX or 40XX AFSCs are now the only logistics AFSCs UPT eliminees are permitted to enter. But even though the number of Academy graduates within logistics career areas is of major concern, an equally important topic is the attitudes and opinions of those Academy graduates already in a logistics AFSC. Most Academy graduates (81%) said that the amount of logistics career field information provided at the Academy was not sufficient for entry into logistics. Many also said that they received either "no logistics career field information" (38%) or "a little logistics career field information" (44%). Of those indicating some information had been provided, they identified CONUS field trips, instructors, professional military studies and training, and Career Day as the primary sources. However, the largest group of respondents (31%) believe they were at least fairly well prepared by the Academy for their current logistics assignment. Moreover, most (54%) believe they were somewhat better prepared than their OTS and ROTC counterparts due to leadership and military training at the Academy (Roberts, 1989).

In gauging how well the Academy is portraying logistics assignments to Academy cadets, about half (55%) said they did not perceive logistics assignments favorably while at the Academy and 81% said, as a cadet, they perceived a logistics assignment as a "second rate job." Possibly as a partial result of this, a logistics AFSC was not the first choice of assignment for 72% of these graduates. Additionally, most respondents believe personnel from non-logistics AFSCs also perceive logistics as a poor AFSC in which to work. Still, most Academy graduates in logistics changed their perspectives once in the career area. Most are satisfied with their present assignment (84%). While they believe promotion opportunities to either Colonel (34%) or General officer rank (82%) are below average or poor, it is fair to excellent (85%) up to this point. Strangely enough, while 43% believe their promotion opportunities are worse than those throughout the rest of the Air Force, 66% say they would not improve their situation by being in other non-rated AFSCs. Perhaps this is because over half would not transfer to another AFSC at this point in their careers. Eighty-six percent believe that either the Aircraft Maintenance and Munitions or the Acquisition Contracting/Manufacturing career areas are the most desirable even though most of these respondents are in other logistics AFSCs. Ending on a positive note, 57% indicated their current working conditions are good or very good. These results did not significantly differ with respect to grade or time in service.

Discussion

The intent of this project was to identify current force structure characteristics and collect attitude and opinion data from USAFA graduates in logistics AFSCs. It was not to evaluate Air Force policy on Academy graduate assignments even though we have identified that present policy affects Academy graduate representation. However, it may be appropriate to look at the maximum direct duty assignment quotas from the Academy at some future point. Unfavorable attitudes held as cadets may also be responsible for low representation. Survey results indicated there would be many volunteers from various base/wing level logistics organizations willing to discuss the many benefits, rewards and challenges of becoming a logistician to Academy cadets during Career Day or other formal programs. And even though graduates do not believe they were well prepared for the technical aspects of a logistics career, it was beyond the scope of this project to objectively measure respondents' job performance and prove this one way or another. It is interesting to note that while graduates did not have a positive attitude towards logistics as cadets, these opinions changed after experiencing the challenges and rewards of a logistics assignment.

References

Boles, Billy J. (1989). Initial assignment of USAF academy graduates: Reply Letter. HQ AFMPC/CC.

Hamm, Charles R. (1989). Initial assignment of USAF academy graduates: Reply letter. HQ USAFA/CC.

Hickey, Thomas J. (1988). Initial assignment of USAF academy graduates: Action memorandum. HQ USAF/DCS.

Roberts, William W. (1989). Source of commission by specialty: DMDC report from 8812 active duty master file [Machine-readable data file]. DMDC/DPMYL (Producer). DMDC/DPMYL (Distributor).

Roberts, William W. (1989). [USAF academy graduates in logistics: Survey and results]. Unpublished raw data.

The Situational Awareness Component of Cockpit Resource Management

Joseph L. Bunecke, Major, USAF Academy
H. Kingsley Povenmire, University of Dayton Research Institute
Marty R. Rockway, University of Dayton Research Institute
Mark W. Patton, Captain, USAF

Abstract

This paper establishes the correlation between B-52 Weapon System Trainer (WST) performance and one Cockpit Resource Management (CRM) tool, the LINE/MOST Worksheet (for a summary see Helmreich & Wilhelm, 1989). Through further examination of this relationship, the authors recommend specific behaviors to enhance CRM skills and identify an informational basis for situational awareness (SA).

The primary goal of CRM research is to improve crew effectiveness, but regardless of the quality of interrelational skills a crew exhibits, task accomplishment requires a solid informational base. Many researchers have defined, and to some degree, quantified the composition of the data necessary for a pilot to effectively perform a mission (for a review, see Fracker, 1988). These attempts to operationalize the term situational awareness have centered on the perspective of the individual pilot. For a crewed aircraft however, these efforts are certainly applicable to the pilot-in-command since he bears the ultimate responsibility for the safe conduct of the flight, yet insufficient for the crew as a whole. All crew members must maintain an awareness of the dynamic relationships between the aircraft and several rapidly-changing variables and employ CRM skills to accurately assemble and utilize this knowledge.

Several investigators have identified information transfer as a group process variable indicative of effective performance. According to Foushee & Helmreich (1988), information flow is "one of the most significant" (p. 209) predictors of performance, and Lanzatta & Roby (1960) state that "the way a group 'utilizes' its resources and the procedures it employs for communicating essential information are as important, if not more important than 'knowledge' of the problem for determining its performance" (p. 146). In an operational Strategic Air Command (SAC) B-52, efficient and effective information transfer is imperative due to the physical separation of the aviation, offensive, and defensive crew positions. The fact that the aircraft's interphone system is the primary medium through which a B-52 crew maintains situational awareness affords a unique research opportunity. Since crew members verbalize their understanding of the situation, it seems that an analysis of interphone communications might provide (1) an assessment of CRM skills and (2) evidence in support of the theoretical content of the situational awareness construct.

Method

Subjects

Seven operational B-52 crews served as subjects. In an effort to ensure a wide range of performance, the authors selected crews following procedures suggested by Helmreich, Foushee, Benson & Russini (1985) and Ginnett (1987). Squadron Operations Officers ranked crews on the basis of overall mission capability and one experimenter chose crews from the top and bottom thirds. The other experimenters were not aware of overall crew ranking.

Procedure and Apparatus

When each crew arrived at the WST facility, the experimenters introduced themselves, explained the purpose of the research and stressed the anonymous nature of the data collection. Representatives from the First Combat Evaluation Group (1CEVG), the SAC-wide standardization and evaluation division, presented a realistic, short-notice mission briefing, and allowed the crews 15-20 minutes to review the scenario before proceeding to the WST. Crews flew a simulated combat mission in the B-52 WST which has high-fidelity crew positions for all six crew members: pilot (P) and copilot (C) in the aviation module; radar navigator (R) and navigator (N) in the offensive module; and the electronic warfare officer (E) and gunner (G) in the defensive module. It is equipped with an out-the-window visual system in the aviation module and provides an extensive capability for simulating a wide range of threats, malfunctions, and environmental conditions. The 1CEVG personnel developed a realistic, high workload, wartime, mission scenario for use in this study. The mission involved takeoff, flight through various threats, a malfunction necessitating crew interaction, attack of a heavily defended enemy target, and navigation to friendly territory. The scenario incorporated the characteristics and features of Mission-Oriented Simulator Training and crews flew the entire profile without instructor intervention. During each simulator sortie, three highly-trained CRM observers manually recorded all communications. The 1CEVG observers and the local simulator instructors ranked crews on mission performance as defined by threat avoidance and bombing accuracy. After the simulator sortie, the 1CEVG personnel debriefed crew members on their performance and CRM observers rated each crew using the LINE/MOST Worksheet. After compiling transcripts, the authors evaluated the various aspects of crew interaction.

Results

Figure 1 depicts a positive correlation between 1CEVG Performance Ranking and LINE/MOST ranking ($r = .835$, $p < .05$).

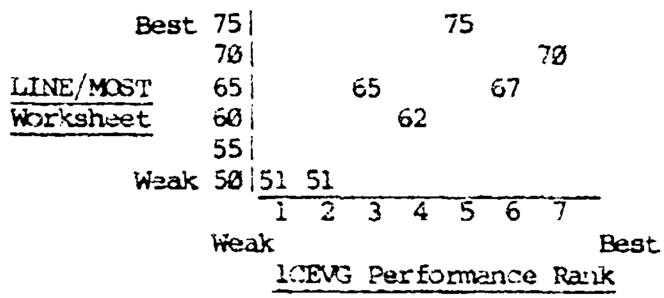


Figure 1. LINE/MOST Score versus 1CEVG Performance Rank

This result identifies a relationship between performance and the LINE/MOST Worksheet CRM assessment tool. Table 1 shows significant correlations between individual LINE/MOST items and the total LINE/MOST score. Review and analysis of the transcripts centered on LINE/MOST Worksheet items which correlated strongly with total LINE/MOST score. These elements seemed influential in improved CRM performance and indicate categories of interphone exchanges which would most likely contain evidence concerning aspects of SA.

Table 1. LINE/MOST Score versus Individual LINE/MOST Items

SKILLS		
4.	Inquiry practiced.	.81*
5.	Advocacy practiced.	.75*
10.	Distractions avoided or priortized.	.91**
11.	Workload distributed and communicated.	.86*
17.	Conflict resolution.	.91**

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

Discussion

Bolman (1979) refers to SA as the crew's Theory of the Situation. He highlights two skills which influence the probability that a crew will detect and revise a faulty theory: (1) a Theory of Practice which builds in inquiry and testing in anomalous situations, and (2) abilities of crewmembers to combine skills in advocacy and inquiry. Although they are not the variables with the strongest correlation to overall LINE/MOST scores, instances of inquiry and advocacy in the mission transcripts readily lend themselves to analysis.

It is useful to think of advocacy as any instance in which a crewmember endorses a particular course of action. A crewmember demonstrates inquiry whenever he realizes that information is lacking or that incongruencies exist in the crew's Theory of the Situation and he conveys this uncertainty via question or statement. In this scenario, the number of turns required in the route characterized N to P transmissions; the amount, type and location of threats constrained E to P exchanges; and how well the pilot remembered and carried out these directions influenced his communications with N and E. An analysis of the context in which advocacy and inquiry occurred yielded the following specific behavioral recommendations to enhance CRM skills. The implications for the content of an SA definition follow each recommendation.

- <u>Keep crew advised of position-specific information.</u> Some information is available to only one crew member/position. In the crews that performed best, individuals made their information base available to the entire crew. The P/C should keep the crew advised of pertinent flight parameters; the R/N should transmit the relationship of the aircraft to the terrain; and the E/G should verbalize the aircraft's relationship to ground/airborne threats. Sometimes crew members take easily accessible information for granted. When each member tries to convey a complete picture of his understanding of the situation, the entire crew benefits. It seems that the more successful crews had a greater quantity of interphone information available for processing.

<u>Corollaries</u>

— Listen for interphone transmissions pertinent to job performance.

— If information essential to performance isn't available, ask for it.

— Maintain intercom discipline, i.e., don't interrupt other transmissions.

<u>Situational Awareness Implications:</u> The pilot must be aware of the aircraft's vector in the three-dimensional volume of air (i.e., heading, altitude, and airspeed), the location of this vector with respect to the terrain over time (i.e., navigation), and the aircraft's relationship to any obstacle which might affect his ability to aviate and achieve the mission's goal (i.e., air/ground threats, weather, equipment malfunction, etc.). Crew members must be aware of this information to the extent that they must use it to perform their duties.

- **Confirm information if possible; challenge if doubtful.** This recommendation is similar to inquiry and advocacy, but it goes even further by requiring an increased level of confidence. By corroborating or questioning interphone transmissions, the crew either validates and reinforces information or focuses attention on reassessment.

Corollaries

— Verbalize confirmations and challenges. A "thumbs-up" or shoulder shrug limits information dissemination and precludes scrutiny by other crew members.

— Resolve conflicts.

Situational Awareness Implications: When information incongruities exist (i.e., conflicting crew member recommendations or instrument indications), interpret them as clues that SA might be incomplete.

- **Provide information to increase the range of options; be flexible.** The pilot must process and act upon an extremely large amount of information in a minimal amount of time. Sometimes he must combine interphone information with knowledge that has just become available to him. Advocating and sticking to only one course of action needlessly limits the possible solutions to a problem. Due to the rapidly changing nature inherent in aviation tasks, providing information to increase flexibility might preclude needless questioning if the situation warrants an immediate change in tactics/strategy.

Corollary

— Communicate all pertinent information and a confidence level if possible. Even if erring on the safe side, filtering or interpreting distorts information.

Situational Awareness Implications: SA is an extremely dynamic understanding of complex relationships.

- **Share workload within cockpit/station.** Sometimes crew members queried each other because they were just too busy to pay attention to the interphone. The information was available, but not perceived and/or understood. The good crews managed to minimize these extraneous transmissions by sharing the workload between the two individuals at each position. Vigilance studies have shown that human beings are not proficient monitors. Including crew members in the task not only keeps them "in the loop", it also frees hands/resources to devote to other tasks. Although it seems advisable for all stations to share the load, it's almost imperative that the P relinquish some responsibilities to the C in order to accomplish the primary task: maintaining aircraft control.

Corollaries

— Use all equipment/resources available (i.e., autopilot, etc.).

— Do your job well first, but try to help others do their job IF your workload permits.

Situational Awareness Implications: There are limits to human information processing capabilities. Predict periods of high workload, plan accordingly, and anticipate consequences of actions.

Conclusion

The behavioral recommendations above not only function to improve crew effectiveness, but also provide an indication of SA components in general. They focus on the information processing aspect of aircrew performance while CRM courses develop group skills to insure the efficient use of this knowledge. As this research demonstrates, study of the B-52 aircrew and its unique mission can contribute to the understanding and improvement of both components of aircrew performance.

References

Bolman, L. (1979). Aviation accidents and the 'theory of the situation.' In G. E. Cooper, M. D. White, & J. K. Lauber (Eds.), Resource Management on the Flight Deck: Proceedings, NASA/Industry Workshop (NASA CP 2120 pp. 31-58). Moffett Field, CA: NASA.

Foushee, H. C. and Helmreich, R.L. (1988). In E. L. Weiner and D. C. Nagel (Eds.), Human Factors in Aviation. San Diego: Academic Press.

Fracker, M. L. (1988). A Theory of Situational Awareness: Implications for Measuring Situational Awareness. In Proceedings of the Human Factors Society 32nd Annual Meeting (pp. 102-106). Santa Monica, CA: The Human Factors Society.

Ginnett, R. C. (1987). First encounters of the close kind: The formation process of airline flight crews. Unpublished doctoral dissertation, Yale University, New Haven, CT.

Helmreich, R. L. & Wilhelm, J. A. (1989). Reinforcing and measuring flight crew resource management: Training captain/check airman/instructor reference manual (NASA/UT-IM-87-1). Austin, TX: Department of Psychology, The University of Texas at Austin.

Helmreich, R. L., Foushee, H. C., Benson, R. & Russini, W. (1985). Cockpit resource management: Exploring the attitude-performance linkage. In R. S. Jensen & J. Adrion (Eds.), Proceedings of the Third Symposium on Aviation Psychology, (pp. 445-450). Columbus, OH: The Ohio State University, Aviation Psychology Laboratory.

Lanzatta, J. T. & Roby, T. B. (1960). The relationship between certain group process variables and group problem-solving efficiency. Journal of Social Psychology, 52, 135-148.

Aspects of Workload For Military and Civilian Decisionmakers

Elliot E. Entin
ALPHATECH, Inc.

Abstract

Weights ascribed to three workload dimensions of a task thought up in abstraction by civilian decisionmakers as part of the scale building phase of the Subjective Workload Assessment Technique (SWAT) were not congruent with weights they assigned to the dimensions when performing a decisionmaking task. This incongruency was not found for military-experienced subjects. The findings may indicate a flaw in the scale building procedure of the SWAT.

Recently stress has become an important topic for practitioners and researchers. Various unfortunate events around the globe have pointedly indicated that stress can have a marked impact on human performance, and often this effect is seriously deleterious. In many instances the primary stressors were identified as workload related. That is, individuals were in situations where there were too many tasks to complete, there was insufficient time available to complete the required tasks, or the mental effort needed to perform the required tasks was very high. These events have, in turn, evoked a renewed interest among researchers in finding a way to characterize and assess workload.

Unfortunately, no universally accepted definition, characterization, or assessment of workload exists (Lysaght et al., 1989). These researchers note that workload definitions are often discussed in the following three broad categories: amount of work and number of things to do, time and the particular aspect of time one is concerned with, and the subjective psychological experiences of the human operator. Of these, we are particularly interested in the subjective psychological characterization in our examination of stress and workload. That is, we wish to focus on the individual's subjective internal experience when performing a particular task. A statement by Johanssen et al. (1979) sums up our position well, "If the person feels loaded and effortful, he is loaded and effortful whatever the behavioral and performance measures show" (p.105). We too believe workload must be assessed subjectively. To this end we have employed the workload assessment method devised by Reid et al. (1981) called the Subjective Workload Assessment Technique (SWAT).

The SWAT specifically characterizes workload according to three dimensions: time constraints, task complexity, and psychological stress. The assessment process involves two distinct phases. Phase one is carried out prior to data collection. Each participant performs a card sort as a precursor to development of a workload scale. Each card contains a different combination of the three workload dimensions: time, mental effort, and psychological stress. In turn, each workload dimension is described in terms of three levels: low, moderate, and high. Thus, participants sort a set of 27 cards (3 levels of time x 3 levels of mental effort x 3 levels of psychological stress) so that the cards are rank-ordered according to the level of workload described. The rank-ordering is then subjected to conjoint measurement analysis to determine the relative weights of each dimension and to establish a scale unique to each individual.

The instructions to the participants specifically indicate that they should think of a generic task while doing the card sort and weight the workload descriptors on a card as to the amount of workload each imposes for such a task. (Obviously, there is no way of knowing what kind of task each person has conjured up in his mind).

The second phase occurs during data collection. At the end of a job, trial, or experimental condition, participants rate the workload they just experienced according to the three dimensions of time, mental effort, and psychological stress. Ratings are done on a three-point scale (low, moderate, and high) for each dimension. (Note, in this case subjects are rating their workload experience for a specific concrete task.) Using software provided by Reid et al. (1981), each SWAT score which consists of one number from each of the dimensions (e.g., 1, 2, 2), is converted to a percent workload score on the participant's unique scale. Zero percent represents very low workload, while 100 percent represents very high workload.

Overall, the scale has shown good reliability and several indications as to the SWAT's validity have appeared in the literature (Lysaght et al., 1989). In our work we have used SWAT almost exclusively to assess subjects' workload experience performing decisionmaking and discrimination tasks. In some instances we have omitted the card sort and scale building phase and devised a rating scale based simply on the three dimensions (see MacMillan et al., 1987 and Entin et al., 1988). This procedure yielded very acceptable results and started us wondering about the efficacy and importance of the card sort and conjoint measurement scale building aspects of the SWAT. We detect an apparent weakness in the SWAT and this weakness is highlighted in the parenthetical expression in the above SWAT description. Subjects are ask to perform the card sort thinking about some generic task and workload in an abstraction. Who knows what kind of generic task subjects conjure up or what repertoire of tasks are immediately available to them? In such an abstraction subjects could look at two card descriptions (e.g., card one: high time load, moderate mental effort, moderate stress load and card two: moderate time load, moderate mental effort, high stress load) and for the particular task they have in mind be very sensitive to stress but relatively insensitive to time and thus rank card one much lower in workload than card two. Had they thought of some other task, the rankings could have been reversed. The task they encounter during an experiment or work period may be different still, and engender yet another set of rankings.

Clearly, phase one of SWAT attempts to derive the subject's generic feelings about the three workload dimension to establish a workload scale that uniquely describes the subject's perceptions. It appears, however, that the SWAT methodology may be flawed. To the test this we derived a workload scale for each of our subjects employing the methodology spelled out in phase one of SWAT. We then compared the weighting of the workload dimensions derived in this manner to the weight derived from a regression analysis of the subjects' workload rating for a specific military-oriented decisionmaking task. We expected to find substantial differences in the two sets of weights, because we expected few subjects to have experience with such a task. We then looked at the same comparisons for a sample of military subjects who have some experience with the specific decisionmaking task. We expected a much higher congruence between the phase one SWAT weights and those derived from the regression analysis. Finally, we compared the workload ratings of the two samples on the decisionmaking task across a number of experimental conditions. We expected the training and experience of the military subjects to lead them to perceive workload differently than the civilian decisionmakers.

Method

Overview of the Experimental Decisionmaking Paradigm

The experimental research paradigm is a modification of the team optimal stopping with communication alternatives paradigm (developed by ALPHATECH) and is an abstraction of an antisubmarine warfare (ASW) detection problem. The subject, assuming the role of an ASW commander, was required to decide if a detection was a hostile or friendly submarine. The decision was based on the pump noise signature of the detected submarine. Decisionmakers knew that the friendly submarine's pump noise signature was always 370 Hz. The hostile submarine had a pump noise signature of 330 Hz and on other occasions a signature of 360 Hz. Decisionmakers also knew that the detection sensors had an inherent error distribution with a standard deviation of 20 Hz. To aid in the discrimination decisionmakers could probe (task their sensors for another measurement) or ask the opinion of a knowledgeable consultant. Information requests always incurred a cost. Stress was induced by imposing a specific deadline (time pressure) and requiring the decisionmakers to attend to an intrusive secondary task. The four within subjects factors: stress (low, high), discriminability (easy, hard), information cost (cost of probe less than, equal to, or greater than consultant), and consultant expertise (equal to, greater than decisionmaker) were nested within two different samples: military and civilian.

Subjects

The civilian sample consisted of nine volunteers from an engineering firm or a college while the military sample consisted of seven officers drawn from the ranks of instructor or student at the Naval War College in Newport, Rhode Island.

Procedure

At the conclusion of each treatment trial subjects completed a SWAT form. Thus a rating for each of the three workload dimensions (time load, mental effort, and stress) was produced for each treatment condition. In addition to the percent workload SWAT score computed for each subject based on his unique scale, a score was also computed for each of the workload dimensions based on the raw rating subjects gave at the conclusion of each treatment trial. It was these latter scores that were subjected to the regression analysis.

Results

The weights for the three workload dimensions derived from the card sort and conjoint measurement analysis of phase one of the SWAT were compared for the two samples. As expected, civilians and military subjects did not weight the workload dimensions the same way. Seventy-five percent of the civilians saw stress as the most important determinant of workload while only 29 percent of the military sample saw stress as most important. Time load was the modal choice of the military sample with 43 percent seeing this dimension as the principle determinant of workload. We surmise that the training and experience of the military subjects has taught them how to cope with stress, thus, they see stress as a less important determinant of workload than the civilians do. Timing, however, is continually emphasized as a crucial factor in military operations and this contributed to the military sample's sensitivity to the importance of time load.

To determine how the two samples weighted the three workload dimensions in response to the ASW discrimination task, a score was derived for each dimension based on the raw ratings collected at the conclusion of each treatment trial. The three dimension scores served as the dependent variables in a stepwise multiple regression analysis to predict the overall workload score (also computed from the raw ratings). For civilians, mental effort was the first variable entered into the regression equation to predict overall workload score and it accounted for 70 percent of the total variance. The next variable entered was time load, which accounted for an additional 18 percent of the variance. Psychological stress only accounted for 12 percent of total variance. As we predicted, the weights attributed to the workload dimensions of a task conjured in abstraction by the civilian decisionmakers did not match the weights assigned when confronted with a concrete decisionmaking task. In the abstract, the civilian sample denoted psychological stress as the most important determinant of workload, but when working on the concrete ASW discrimination task it was mental effort that proved to be the primary predictor of workload. The outcome of the regression analysis for the military sample matched what they had stated in the abstract. That is, time load was the first variable loaded into the regression equation and it accounted for 68 percent of the variance. This was followed by effort, which contributed another 22 percent to the predicted variance. The military sample still differed from the civilians on what was the most important determinant of workload, and appeared to be better calibrated on the weights of the workload dimensions than the civilians. We attribute this calibration to their military training and experience.

Within-subjects multivariate analyses of variance were computed to examine the effects of the independent variables nested within high and low stress conditions. As expected, both samples reported higher workload in the high than low stress situations. This is a kind of manipulation check that the stress induction of time pressure and the necessity to perform an intrusive secondary task did in fact produce higher workload. Moreover, the results also showed that the civilians reported significantly higher workload ratings than the military personnel in both low and high stress conditions. As predicted, military decisionmakers can cope with stress better than the civilians.

Figure 1 shows that civilians report higher workloads in the difficult than easy discrimination conditions for both high ($p<.008$) and low ($p<.003$) stress situations, while no such differences exist for the military sample. Apparently the training and experience enjoyed by the military personnel tends to level out the difference between the easy and difficult discrimination conditions and thus they do not perceive the workload to be greater in the difficult discrimination condition.

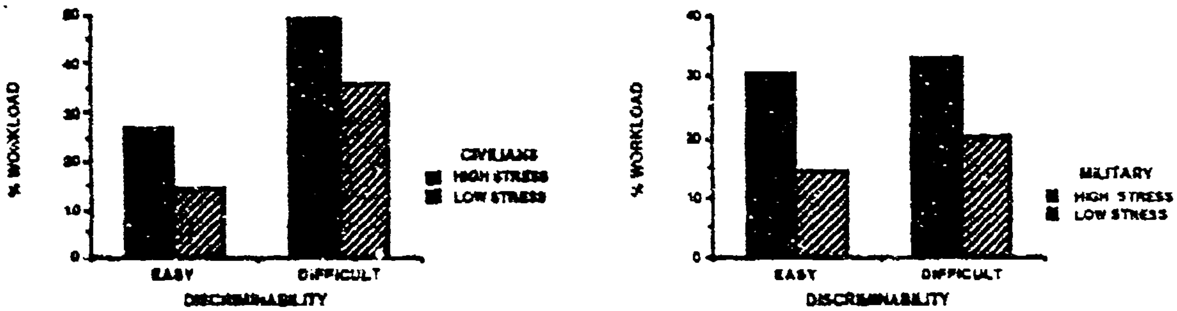


Figure 1: Workload as a Function of Discriminability for Low and High Stress Conditions

Discussion

Our results indicate that the scale building procedures of SWAT may be flawed. The weights ascribed to the three workload dimensions of a task conjured up in abstraction by the civilian decisionmakers were not congruent with the weights they assigned to the dimensions when performing the ASW decisionmaking task. The difficulty arises when subjects are asked to perform the card sort, but are not given an appropriate "target task" to use. Thus, they perform the card ranking for some abstract task they think up. It is very likely that this abstract task has little in common with actual task they will perform, leading subjects to characterize workload in the card sort and in the performance task quite differently.

The results also showed that if subjects' training and experience lead them to think of tasks similiar to the one to be performed, as was the case with the military sample, a much higher agreement will exist between the weights derived from the card sort and those applied to the performance task. This explains why past use of the SWAT, as reported in the literature, has met with good success. The subjects employed in most of the studies (see Lysaght et at., 1989) were highly trained (e.g., test pilots, military personnel) working with tasks well known to them (e.g., cockpit design, simulated air-to-air combat).

We think the scale building procedure of the SWAT is an interesting concept. Our quarrel is with one aspect of the procedure. We recommend that subjects be given the description of a "target task" that is similar to the task they will be performing to refer to while performing the card sort. The workload scales derived for the subjects will now be appropriate for the actual task performed.

We have one further cautionary note. The findings showed that the way civilian and military decisionmakers perceived task workload was different. If researchers wish to study the workload proclivities of military decisionmakers, these results argue that military-experienced personnel should serve as subjects.

References

Entin, E.E., Needalman, A., Mikaelian, D., and Tenney, R. (1988). Experiment II Report: The Effects of Option Planning and Battle Workload on Command and Control, TR-388-1, ALPHATECH, Inc., Burlington, MA.

Johanssen, G., Moray, N., Pew, R., Rasmussen, J., Sanders, A., and Wickens, C. (1979). Final report of experimental psychology group. In N. Moray (Ed.), Mental Workload, Its Theory and Measurement (pp. 101-114), New York: Plenum Press.

Lysaght, R.J., Hill, S.G., Dick, A.D., Plamondon, B.D., Linton, P.M., Wierwille, W.W., Zakland, A.L., Bittner, A.C., and Wherry, R.D. (1989). Operator Workload: Comprehensive review and evaluation of operator workload methodologies, TR-851, U.S. Army Research Institute for the Behaviorial and Social Sciences, Alexandria, VA.

MacMillan, J., Entin, E.E., and Lentz, P. (1987). Experiment Report: The Effects of Option Planning and Battle Workload on Command and Control Effectiveness, TR-368, ALPHATECH, Inc., Burlington, MA.

Reid, G.B., Shingledecker, C.A., and Eggemeier, F.T. (1981). Application of conjoint measurement to workload scale development. In proceedings of the Human Factors Society 25th Annual Meeting (pp. 522-525), Santa Monica, CA: Human Factors Society.

General Issues of Operator Workload[1]

Michelle R. Sams, Ph.D.
Richard E. Christ, Ph.D.
Army Research Institute, Ft Bliss, TX

Abstract

Even with extensive research and development of a myriad of assessment techniques, there are still a number of unsettled and unanswered issues regarding the concept and measurement of workload. This paper discusses the relationship between performance and workload and examines various dynamic, validity, and diagnostic issues of workload.

The Operator Workload (OWL) Program is a just-completed, basic and applied research effort sponsored by the Army Research Institute (ARI). The OWL Program was directed to establish guidance for the assessment of operator workload associated with the operation of Army systems. Throughout the conduct of the program, various general workload issues have surfaced. We are not the first to recognize some of these issues, but feel this is an appropriate venue in which to organize and express our thoughts and those of others.

The Performance Envelope

When evaluating current or proposed systems, analysts may be asked to determine whether task demands exceed the capability of the operators, or in other words, whether workload levels will surpass the "red line". Seven (1989) expressed the need for some criterion to determine what constitutes a "full load", particularly if workload measures are to indicate the "proportion loaded". Some workload techniques are based on such an absolute estimation approach. These measures may define specific values, through their respective algorithms, as the points above which the operator is overloaded and degradation of performance will be exhibited.

However, Hart (1989) points out that while increases in system complexity and mission requirements may result in unreasonably high demands on the operators, automation in some systems may actually reduce workload. In fact, decrements in performance may be evidenced as a result of boredom and complacency. Therefore, it is important not only to determine the level of workload which is excessively high, but also to define the range of optimal or excessively low levels of workload.

It is proposed that human performance is not best conceptualized as the mercury which rises and falls with task demands and measured by a workload thermometer, but rather that human performance be considered in terms of a multidimensional performance envelope (based on Lysaght et al., 1989). In the

[1]The authors wish to acknowledge Allen L. Zaklad and his staff at Analytics for their contributions to the OWL Program and this paper. The views, opinions, and findings contained in this paper are those of the authors and should not be construed as an official Department of Army position, policy or decision.

context of a performance envelope, the "red line" then becomes an outer boundary beyond which performance is either unacceptable or unattainable. This includes situations in which task demands exceed the operator's capability to perform effectively and situations in which task demands are so minimal that performance may be at risk due to boredom and inattention. Within the boundary are a continuum of zones. Zones nearest the boundary represent those situations for which performance is momentarily acceptable but is in danger of becoming unacceptable should certain events occur (e.g., emergencies). The core of the envelope contains the workload comfort zone, where task demands are sufficient to maintain alertness yet do not exert undue pressure.

These boundaries and zones are not static, but are responsive to various internal and external determinants. These include factors of job requirements and constraints as well as the human's capacities and resources. Aspects of these determinants contribute to the variability that is found in workload measurement. The determinants may differ from operator to operator, from day to day, and even from moment to moment. Perhaps more importantly, these determinants may interact. Performance can thus be described as a momentary point in space and time within the performance envelope. The proximity of the individual's position to the boundaries of the envelope is indicative of the operator's relative capacity to respond.

Performance and Workload

Various performance measures have been used to assess operator workload. Seven (1989) proposes that realistic measures of workload are those unobtrusive measures which are specific to the operational system. These real-time measures might include neglected noncritical tasks (e.g., proportion of non-task-relevant messages in air defense systems). Resulting system performance has also been used as an indication of workload (e.g., number of targets successfully engaged). However, many performance measures reflect characteristics of the environment or the system and provide little information about the effort exerted to achieve that performance.

Hart (1989) points out there is a difference between measuring performance and the amount of effort expended to perform the task. Operators may be able to exert the extra effort and allocate the additional resources required to maintain system performance in the face of increasing task demands. However, operators may apply some extra effort but still be unable to meet demands, resulting in system performance degradation. Additionally, operators may elect to maintain a consistent level of effort in the face of increasing task demands, also resulting in decreases in system performance. Finally, operators may simply attempt to achieve acceptable, rather than perfect levels of performance.

It is possible that relationships between performance and workload measures occur only at workload extremes (very high, as well as very low). It may also be that workload scores are more highly correlated with performance for tasks which are critical to system performance (e.g., acquisition and tracking tasks in air defense systems). For some situations, no relationship between workload and performance may be found and, in fact, instances of high workload were found to result in better performance (e.g., Yeh and Wickens, 1988). However, to conclude that there is no change in workload or that a particular workload measure is not valid if there is no accompanying change in performance, is an erroneous conclusion and runs contrary to common sense.

Dynamic Aspects

Human performance has been described as a momentary point in space and time and as a function of many variables. However, human behavior and workload are both dynamic. Workload may be considered from the standpoint of the effort required for only the current task or tasks being accomplished at a particular moment in time. Momentary workload, in this context, depends not only on the task to be accomplished, but also the current resources of the operator to perform the task. It may be that task performance is more dependent on the accumulated workload at various points in time than on the momentary demands of the task. For example, the probability of an error may be more related to accumulated workload than to momentary workload. Current remaining workload is comprised of the total expected efforts required to complete all currently assigned tasks.

It is likely that operators adopt strategies for responding to workload. In some situations, operators may reschedule or shed tasks so that current remaining workload is spread out more evenly over a period of time. In some cases, the operator must be able to make a rapid transition from a low workload state to a high workload state. Along with this transition the operator may adopt alternative strategies for the increased workload (e.g, concentrating on high priority tasks). Response strategies are probably as dynamic as the changes in workload. However, if high workload continues for an extended period of time, the operator's resources may become depleted resulting in inadequate performance. Even extended periods of low workload, accompanied sporadically with sharp transitions to high workload may diminish the physical and mental resources of operators.

Temporal variables may have significant influences on the workload experienced and on the results obtained from the scales designed to measure it. For example, operators who are asked to rate their overall workload from the beginning of a mission until the present time may be utilizing either an averaging method or an additive interpretation of workload. Workload peaks may have a significant effect on workload ratings (see Iavecchia, 1989, for a discussion of this issue). When obtaining workload information prospectively or retrospectively, the operator or SME can be seen as a source of knowledge that may be elicited through a structured format (e.g., via a workload rating scale). When operators give workload ratings while "on the job" they can be seen as "workload meters" that simply read out a load level (Vidulich, 1989).

Validity

Workload measures which vary in relationship to the manipulated levels of task difficulty (e.g., orders of control for tracking tasks) are said to have demonstrated face validity. It also has been argued that operator ratings are the most valid measure of workload, regardless of performance levels. "If the person feels loaded and effortful, he is loaded and effortful whatever the behavioral and performance measures show" (Johanssen et al., 1979, p. 105). In contrast, Wickens (1989) states that "the ultimate design criterion should be satisfactory performance rather than adequate subjective opinion" and that it is the former that should be used for validation of workload measures (p. 271). And so the debate continues.

As has been previously illustrated, workload is a function of many variables, of which task difficulty is only one. Since different types of measures are likely to be related to different system or mission variables, it is suggested that a battery of operator workload measures will obtain the most useful data. This is particularly important in a field test situation where it is imperative to glean every available bit of information possible. Validity in the field is often a "patchwork" of bits of information, most meaningful in their totality. This approach employs the method of converging operations, a diagnostic methodology for attacking a problem in several different ways to insure the quality of the inference.

The validity of a workload assessment study is also related to the realism and fidelity of the test situation. Workload measures obtained in early field tests, in which many aspects of realistic battlefield situations are absent, most likely underestimate the workload that would occur under more realistic conditions. Therefore, any workload issue(s) identified in these studies would be indicators of more severe problems in situations in which failure has more severe consequences. The general trend is toward utilizing more sophisticated technologies to deliver increasingly realistic environments. With the increasing cost of such technologies, it is important to screen the experimental variables to find out which are responsible for substantial portions of the response variance and hold less readily manipulated variables constant (Roscoe, 1980).

Diagnosticity

The essence of the diagnostic process is to identify the specific mechanism or process involved during the performance of a particular task. Therefore, the diagnostic capabilities of any of the workload techniques must rely on a thorough task analysis. The result of the analysis is an operator activity profile as a function of mission time and segment, essentially a time-based analysis of performance requirements. Without such detail, a workload analysis can only identify general problem areas leaving the determination of the cause(s) of the workload to guesswork.

It is important to note that the diagnosticity of workload techniques lies in their ability to pinpoint the <u>sources</u> of workload, which is not to be confused with the identification of the <u>causes</u> of workload. For example, the OWL Program found that for the IOS-F-H air defense system identifying rotary-wing aircraft was rated higher in workload than for fixed-wing aircraft. The source of workload is a mission variable — the fact that it is a rotary-wing is not the cause. The cause of the higher workload may be that the display does not clearly depict rotary-wing craft, that the unpredictable nature of typical rotary-wing flight paths (e.g., pop-up) makes initial sighting of the target difficult, or other factors. Identification of the source of workload is an important step towards discovering the cause of a potential or demonstrated performance problem. However, it is the determination of the cause which provides the necessary information for steps to be taken to reduce workload (e.g., through changes in human engineering design, training).

Summary

Hart (1989) points out that much confusion has surrounded the term, workload, because it has been used as a label for the stimulus or cause (i.e., objective task demands), the response or the consequence (i.e., effects of such demands on the operators) and as an intervening variable to explain the relationship between task demands and performance. She believes that these various conceptualizations of workload have contributed considerable confusion and apparently inconsistent results to the study of workload.

Even in this short treatise addressing workload issues, the unanswered questions, conflicting results, and opposing viewpoints illustrate this problem of workload definition. Those that question the validity of "subjective" ratings for assessing workload are defining workload in terms of the task demands. The virtues of operator ratings are expounded by those who equate workload with effort expended. If workload is viewed as the "cause", then performance is the "effect" and a monotonic relationship between the two variables is expected. Assumptions of consistently high motivation states of human operators may detour researchers away from studying alternative workload strategies. The research and conclusions drawn about workload are clearly influenced by the researcher's conceptualization of the term.

References

Hart, S. (1989). Workload: A new perspective. Unpublished manuscript.

Iavecchia, H. (1989). The effect of high task load recency on subjective workload reporting. Unpublished manuscript, Villanova University.

Johannsen, G., Moray, N., Pew, R., Rasmussen, J., Sanders, A., and Wickens, C. (1979). Final report of the experimental psychology group. In N. Moray (Ed.), Mental workload, its theory and measurement (pp. 101-114). New York: Plenum Press.

Lysaght, R.J., Hill, S.G., Dick, A.O., Plamondon, B.D., Wherry, R.J., Jr., Zaklad, A.L., and Bittner, A.C., Jr. (1989). Operator workload: Comprehensive review and evaluation of operator workload methodologies (Technical Report No. 851). Alexandria, VA: U.S. Army Research Institute for Behavioral and Social Sciences.

Roscoe, S.N. (1980). Aviation psychology. Ames, Iowa: Iowa State University Press.

Seven, S.A. (1989). Workload measurement reconsidered. Human Factors Society Bulletin, Vol. 32, No.11, 5-7.

Vidulich, M.A. (1989). The use of judgment matrices in subjective workload assessment: The Subjective WORkload Dominance (SWORD) Technique. Proceedings of the Human Factors Society 33rd Annual Meeting. Santa Monica, CA: Human Factors Society, 1406-1410.

Wickens, C.D. (1989). Models of multitask situations. In G. McMillan (Ed.), Applications of human performance models to system design. New York: Plenum Publishing Corporation.

Yeh, Y.Y., and Wickens, C.D. (1988). The dissociation between subjective workload and performance. Human Factors, 30, 111-120.

Effect of Communication Degradation on Military Crew Task Performance

Leslie A. Whitaker, Ph.D.
Klein Associates
Yellow Springs, OH 45387

Leslie J. Peters, Major U.S. Army, and Georges Garinther, P.E.
Human Engineering Laboratory
Auditory Performance Team
Aberdeen Proving Ground, MD

Abstract

Twelve Bradley qualified crews were tested in SIMNET-T Bradley Fighting Vehicle (BFV) simulators under varying levels of speech intelligibility. Their navigation performance over 20 short (1.5 to 2 kilometer) course exercises was measured. The number of exercises successfully completed and the number of check points reached within each exercise decreased as speech clarity was degraded.

Communication is a factor whose importance is acknowledged in all environments requiring information transfer. McCormick (1976) has described the type of message which is best conveyed using auditory communication:

> The message is simple, short, will not be referred to later, deals with events in time, and calls for immediate action (p. 43).

All these characteristics hold for the communication of information during a military operation, which is run off in real time and requires auditory communication if it is to be executed properly. In many large vehicles (including most military track vehicles), auditory communication among the various members of the crew is critical just to allow navigation to reach a military objective.

An experiment was conducted using the SIMNET-T facility to test the effects of degraded auditory communication on navigation tasks.

Method

Exercise scenarios were developed to test Bradley Fighting Vehicle (BFV) crews on both navigation and gunnery exercises. The navigation task consisted of a point-to-point route finding exercise. The Bradley Commander (BC) called each designated check point along the route in to the base station.

Thirty-six Bradley qualified soldiers were tested as 12 three-man crews. Each crew was composed of a driver, a gunner, and a commander. Crews usually had not worked together as BFV crews before this experiment. Each man's MOS qualified him for his respective crew position. All crew members were tested and found to have an H1 hearing profile.

A repeated measures experimental design was used. Each crew was tested at five levels of intelligibility. A chopping circuit was used to set these levels from very low (approximating 0%) to virtually clear communication (approximating 100%). At each level of intelligibility, each crew completed four exercises. Each crew was tested using the MRT to determine their actual intelligibility level at the designated chop circuit setting.

Performance was measured using a combination of time and operational error measures: Success or failure of the mission (operationally: "Did the crew reach the release point within 10 minutes?"); number of check points reached; time to reach checkpoints; and SWAT workload ratings.

Each crew participated in one day of training and two days of testing. Each crew trained on three tasks: Modified Rhymes Test (MRT) calibration, Subjective Workload Assessment Technique (SWAT) prototype, and Simulation Network (SIMNET) exercises. Working as a crew, they read the MRT until they achieved 96% (or better) intelligibility on three consecutive 50 word tests. They completed the SWAT prototyping card sort. They trained for two to three hours on the SIMNET, first doing point-to-point navigation on a road course set by the BattleMaster. Then they completed two sample exercises of the type that were to be used in the experiment. Each crew completed as many repetitions of these exercises as possible in the time available.

Pairs of crews were tested together. Each crew completed four exercises at each of the five levels of intelligibility. The first level tested was always a moderate one. The remaining four levels were counterbalanced for order. Each intelligibility level required approximately a one-and-a-half (1-1/2) hour session to complete. In this time, each crew completed an initial MRT to obtain a pre-test measure of actual intelligibility. Each crew then completed the four exercises. At the end of the session, they completed a post-test MRT.

In this experiment, eight separate courses were set. By starting at each end of the course, 16 courses were created. Each exercise began at a Start Point, ran through three specified checkpoints, and ended at a Release Point. For an exercise, the two crews began at opposite ends of a single course. The total length of the course was 1.5 to 2.0 kilometers. Crews were allowed five minutes to reach the first checkpoint and 10 minutes to reach the release point.

Results

Although attempts were made to insure that each crew completed exercises at each level of intelligibility, for some crews intelligibility could not be dropped to the lowest levels due to difficulty with the communication system. Therefore, more exercises were run at the higher levels of intelligibility. To determine the actual level of intelligibility for a session, the crew's pre-test and post-test MRT scores were averaged. This average intelligibility was used to classify an exercise session into the following five ranges: (0 to 9%; 10 to 34%; 35 to 64%; 65 to 89%; and 90 to 100%). The means of these five ranges were 1.9%, 27.8%, 51.1%, 72.7%, and 96.0%. Respectively, 36, 36, 68, 48, and 52 exercises were run at these five levels of actual intelligibility.

Navigational accuracy and course completion time as well as SWAT ratings are presented in this paper.

Course Completion Time and Operational Errors

The total distance to be traveled was less than two kilometers. On 87 of the 240 exercises (36%), the crews reached the Release Point. For slightly more (96 or 40%) of the exercises, the experimenters stopped the crew because 10 minutes had passed. For 41 (17%) of the exercises, the crew ended the session. Usually this occurred in the lowest intelligibility condition, because the commander was not able to communicate with his driver.

Intelligibility level influenced the number of exercises in which a crew reached the release point. This effect was roughly linear, raising from approximately 10% success rate at the lowest intelligibility level to 60% at the highest. See Table 1, Percent RP, for more exact values.

On exercises which ended successfully (i.e., the crew reached the release point), the mean time did not differ systematically with the intelligibility level (see Table 1, Mean Latency). This latency value is included for completeness. It is difficult to interpret because latencies were available only for exercises on which the RP was reached. More crews were able to complete more exercises as intelligibility improved. Not all crews exhibited the same level of skill. More expert crews were sometime able to reach the RP even under poor communication conditions; therefore, the latencies of less expert crews were not included in the mean latencies at low intelligibility, but were included at higher levels of intelligibility.

Table 1. Percentage of exercises which ended successfully at the Release Point (Percent RP) are listed as a function of intelligibility level (Intel%). The mean latencies for these successful trials are also listed.

Intel%	Percent RP	Mean Latency (sec)
1.9	11	487
27.8	25	416
51.1	35	483
72.7	42	486
96.0	61	408

In each exercise, the crews traveled a fixed route, stopping at three checkpoints prior to reaching the release point (RP). The number of exercises for which crews actually reached each checkpoint increased with intelligibility level. The percent exercises for which each checkpoint was reached are shown in Table 2. However, the amount of time required to reach a checkpoint did not differ systematically with intelligibility. (Note that the time to arrive is only for those exercises on which the checkpoint was reached. That is, success on reaching the check point serves as a gate for measuring the time variable.)

Subjective Workload Assessment

Crews rated their mental workload using the SWAT scale. This measure consists of three subscales to measure mental effort, time pressure, and stress. Each subscale allows low (1), medium (2), and high (3) ratings. Therefore, the minimum summed SWAT rating is 3.0 and the maximum is 9.0.

Table 2. Percent of exercises on which each checkpoint was reached. Data are tabled as a function of Intelligibility level.

	Intelligibility Level (%)				
Check Point	1.9	27.8	51.1	72.7	96.0
1	17	56	66	83	75
2	11	33	54	58	75
3	11	25	44	56	71
RP	11	25	35	42	61

In an earlier study of tank crews, SWAT ratings have proven to be a leading indicator of task difficulty when intelligibility levels were varied (Whitaker, Peters, & Garinther, 1989). In the current study, crews were asked to provide SWAT ratings after each exercise and after each MRT list.

SWAT ratings can be analyzed in two ways. By using a conjoint measurement technique developed by Nygren (1982), the ratings can be transformed to interval data. These interval data may be reported as a general scale for the entire sample, or may be subdivided to be specific to each of the SWAT subscales. Alternatively, SWAT scores can be reported as untransformed ratings for each subscale or for the sum of the subscale scores. Whitaker, Oatman, and Shank (1986) and Biers (1987) have reported high correlations between the transformed data and the rating sum data.

After both the Pre-session (Pre-test) MRTs and the Post-session (Post-test) MRTs, SWAT ratings were obtained from each crew member. Crew members were asked to rate how difficult it was to do the listening part of the MRT. These ratings were tabulated for each SWAT subscale (time, effort, and stress). These ratings were summed to obtain a Sum SWAT rating, which are listed in Table 3 as average crew ratings at each intelligibility level. These ratings indicate that the crews reported a decrease in difficulty as the intelligibility improved from 1.9% to 96%.

After each exercise in the session, each crew member provided a SWAT rating of the difficulty of the exercise. Task workload (Sum SWAT ratings) decreased as intelligibility improved. The crew's Sum SWAT ratings are also listed in Table 3.

Table 3: Average crew Sum SWAT ratings by Intelligibility level.

Intel%	MRT Pre-Test SWAT Sum	MRT Post-Test SWAT Sum	Exercise SWAT Sum
1.9	7.1	7.5	6.8
27.8	5.9	5.8	5.9
51.1	4.8	4.5	4.3
72.7	4.4	4.4	4.1
96.0	3.1	3.0	3.9

Discussion

The following conclusions are based on the descriptive data from this sample. There have been no statistical tests conducted on these data to allow inferences to the population of BFV crews.

These results are consistent with the hypothesis that military effectiveness is adversely affected by degraded communication. Crews were less likely to complete the mission successfully when speech intelligibility is degraded. They were less likely to complete each phase line of the mission. Based upon the results of this study, SWAT was shown to be a sensitive measure of subjective workload for these military exercises.

The mean latencies were calculated across all crews. These did not differ as a function of intelligibility; however, the interpretation of this latency measure is difficult because it was obtained from different crews at different intelligibility levels. For example, only one crew was able to complete exercises at the 1.9% intelligibility level. However, 11 crews completed some exercises at the 96% intelligibility level. Crew by crew examination of the latencies does show large individual differences; however, analysis of the crew by crew latency data would not seem to provide very meaningful interpretation because of the missing data (i.e., intelligibility levels at which a crew did not ever reach the release point successfully.)

Further research is planned to study the effects of intelligibility on military performance. One area of interest is the use of communication difficulty as an index of training proficiency. For the current study, all crew members were Bradley-qualified in their MOS, yet crews differed widely in their performance under conditions of degraded communication. More skilled, better trained, and more proficient crews may continue to perform well under degraded communication conditions. An index of performance level in simulated battle conditions under varying conditions of auditory workload would provide a valuable tool to assess training and skill levels.

References

Biers, D., & Masline, P. (1987). Alternative approaches to analyzing SWAT data. Proceedings of the Mid-Central Ergonomics/Human Factors Society. Urbana-Champaign, Il.

McCormick, E. J. (1976). Human factors in engineering and design. McGraw-Hill: New York.

Nygren, T. E. (1982). Conjoint measurement and conjoint scaling: A user's guide. AFAMRL-TR-82-22. Wright Paterson APB, OH.

Whitaker, L. A., Oatman, L. C., & Shank, M. D. (1987). Measuring mental workload: A performance battery. HEL Technical Report (TM 21-87). MD: Aberdeen Proving Ground.

Whitaker, L. A., Peters, L., & Garinther, G. (1989). Tank crew performance: Effects of speech intelligibility on target acquisition and subjective workload assessment. Proceedings of the Human Factors Society 32nd Annual Meeting.

New Zealand Military Family Values

Edna J. Hunter, Ph.D.
United States International University
Carol Jarman
New Zealand Defence Psychology Unit

Abstract

During the 1988-1989 Fulbright award cycle, a landmark study of New Zealand military families was conducted by this writer while serving as a Senior Scholar under a 6-month NZ-US Educational Foundation grant. The purpose of the study was to explore major issues/problems of NZ military families and make recommendations to the NZ Ministry of Defence, based on the findings. The focus of this paper is the differences which exist between NZ military family value orientations and families within the American military system.

Although much has been written on military families in the United States (Hunter, 1982), almost no research has been undertaken in New Zealand (NZ). This report will address some of the findings concerning basic NZ military family value orientations, and compare them to what has previously been found for American military families.

Method

The 1988 New Zealand Military Family Study (Hunter, 1989) was the first major study ever conducted of that particular population. This writer was invited by the NZ Secretary of Defence in 1986 to carry out such a study. It was jointly sponsored by the NZ-US Educational Foundation (Fulbright Program) and the NZ Ministry of Defence, under the aegis of the NZ Defence Psychology Unit. Military planners were primarily interested in obtaining recommendations for changes or additions in policy and programs which could perhaps alleviate family stresses and thus make for higher quality of life for NZ military families. The present report will not focus on those recommendations. Rather, it focuses on only a small portion of the larger study, specifically NZ military family value orientations. Participants included spouses (legal and recognized de facto) of all personnel serving in the NZ Armed Forces in 1988 (Army, Navy, and Air Force). The study was bipartite: (1) a mailed paper and pencil survey, and (2) personal interviews with a small representative sample of the total population of spouses. Participation was voluntary.

---

[1] The authors extend special thanks to LCOL Janet Turner and MAJ Alan McCone for their assistance with the project. The views expressed in this report are those of the authors, and do not necessarily reflect those of sponsoring organizations.

A total of 2772 spouses returned the mailed questionnaire, a 51.2% return rate. Personal interviews were conducted at military locations throughout New Zealand with 153 of the spouses who had previously completed the mailed survey. All interviews were taped. For the most part, selection of interviewees was random, although an attempt was made to match, by group, the total population of military families with respect to age, working versus non-working spouses, and presence or absence of children. Results from the mailed survey are discussed elsewhere (Hunter, 1989). This report focuses only on results from the 2-hour semi-structured interviews. A small comparison group (N = 29) of civilian NZ spouses was also selected. The two NZ subsamples of families (153 military, 29 civilian) were well-matched with respect to age range, presence or absence of children, employed versus non-employed, and percent male spouses. As part of the interview protocol, each interviewee was asked to complete a standardized instrument which provides a profile of the relative importance of ten family values or orientations, the Family Environment Scale (Moos & Moos, 1986). The value orientations measured by the Scale include: Cohesion, Expressiveness, Conflict, Independence, Achievement Orientation, Inter-Cultural Orientation, Active-Recreational Orientation, Moral-Religious Emphasis, Organization, and Control.

Findings

Within the group of 153 spouses who were interviewed, 27.5% were spouses of Officer personnel and 72.5% were of Other Ranks (comparable to U.S. enlisted and non-commissioned officers). Most were wives (95.1%); 4.9% were husbands of military women. Their ages ranged from 21 to 58 years of age, with a mean of 31.1 years (S.D. = 6.2). Slightly over twelve percent (12.1%) were dual military couples (compared to 9.3% in the survey sample). Out of that twelve percent, 5.0% were Army, 4.3% Air Force, 2.1% Navy, and .7% Territorials (Reserves). As to service affiliation, 45.8% of those spouses interviewed were Army spouses, 38.0% Air Force, and 15.5% Navy spouses. Over sixty percent (62.0%) were employed outside the home, and over eighty percent (82.9%) had children. Both the survey sample and the interview sample were fairly representative of the total population of married service personnel in the NZ Forces, except for two minor differences. A larger percent of the interview sample had children, compared with the survey sample. Also, dual military couples were intentionally over-sampled in the interviewing in order to obtain as much information as possible about that important subsample of military families.

Findings from the Interviews

Only small portions of the data from the interviews are presented in this paper. Additional findings will be presented at a later time. At some point during each interview, interviewees were asked, what the "best" thing about military life was, and what was the "worst"? As for the "best", they listed (1) financial security, (2) social life, (3) the unique lifestyle with opportunities for travel, (4) affordable housing, and (5) the fact that their military spouses enjoyed it. In response to the question concerning the "worst" aspect of the military lifestyle, they mentioned (1) family separations, (2) mobility, (3) lack of information, (4) poor housing, and (5) lack of privacy. They were also questioned as to what changes or additions they would like to see made to make military family life more satisfying or less stressful. Among the most frequently listed changes/additions, in descending order of frequency of mention were: (1) more information available to family members, (2) spousal input regarding future postings of the military member, (3) more family-related activities, (4) better communications between the military spouse and the family during separations due to military duties, (5) better communication channels between the family and the military unit, (6) longer postings, and (7) confirmed transportation benefits for family members.

Scores on the Family Environment Scales

Between-group comparisons for FES Subscale scores were made between (1) NZ military spouses and NZ civilian spouses; (2) NZ military and American military; and (3) FES Norms and both NZ military and American military. The FES gives a profile of the family's orientation or emphasis placed on various values by family members. There is no overall score, but rather a profile for the individual on the 10 Subscales.

<u>NZ Military versus NZ Civilian Families on the FES</u>. Comparing NZ military spouses with NZ civilian spouses, it was found that there were a number of statistically significant between-group differences on the FES Subscales. NZ military spouses scored significantly <u>higher</u> on Achievement Orientation and Control, and <u>lower</u> on Cohesion and Independence, compared to NZ civilian spouses. There were no statistically significant differences between the groups for the other Subscales. A number of comparisons for family environment for NZ families <u>in general</u> (that is, military and civilian groups combined, $N = 182$) were made to determine differences between: (1) employed NZ spouses versus non-employed spouses; (2) younger versus old NZ spouses; and (3) NZ wives versus NZ husbands. Employed spouses differed from non-employed spouses on Cohesion and Independence, with the employed group scoring significantly <u>higher</u> on both ($p < .01$). Female spouses did not differ significantly from husbands on any of the FES Subscales. There were some differences with respect to age. The younger group (25 years and under) scored significantly <u>lower</u> on FES Organization and Cohesion ($p < .01$). Thus it appears that as spouses grow older, they tend to increase their organizational skills, along with developing higher levels of

cohesion within the family.

NZ Military versus American Military Families' FES Scores. The reader may wonder how NZ military families' FES scores compare with those of military families in the United States. Since the FES has been widely used in America, it is possible to give a tentative answer to this question. Contrasting scores on the FES for the 153 NZ military families in this study with a sample of 71 American military families (data from a study by Nice, McDonald, & McMillian, 1981), it was found that NZ military family members scored higher on Cohesion, Expressiveness, Intellectual-Cultural Orientation, Active-Recreational Orientation, and Organization, and lower on Conflict, Independence, Achievement Orientation, Moral-Religious Emphasis, and Control, with the greatest differences between the two groups on Control and Achievement Orientation (NZ lower on both). No statistical tests of the significance of these differences were possible since raw data for the American sample were not available.

NZ and American Military FES Scores versus Scale Norms. The question also arises as to how both military groups (NZ and American) stack up against FES Norms. When NZ military families' FES Subscale scores were compared with those shown in the Scoring Manual for the sample of 1125 families used to establish Norms for the FES, NZ military families scored higher on Cohesion, Expressiveness, Intellectual-Cultural Orientation, Active-Recreational Orientation, and Organization, and lower on Conflict, Independence, Achievement Orientation, Moral-Religious Emphasis, and Control. Again, no statistical tests of significance were possible. Comparing the previously mentioned group of 71 American military families (Nice, McDonald, & McMillian, 1981) with scale Norms from the Manual, Americans scored higher on Cohesion, Expressiveness, Independence, Achievement Orientation, Active-Recreational Orientation, Moral-Religious Emphasis, and Control, and lower than the norms on Conflict and Independence. American military spouses had exactly the same mean score for Intellectual-Cultural Orientation as the Norm. No tests of statistical significance were computed.

Discussion

In the earlier study by Nice and associates, those investigators concluded that American military families tend to show a "structure-oriented" profile; that is, they score higher than the Norms on Cohesion and Organization, and lower on Conflict. Both NZ military and U.S. military families scored higher than the Norms on Cohesion, Expressiveness, Active-Recreational Orientation and Organization, and both scored lower on Conflict. Thus, results from the 1988 NZ Military family Study corroborate those of Nice and colleagues. Also, NZ military spouses showed higher scores than either American military families or standardized Norms on Cohesion, Expressiveness, Active-Recreational Orientation, and Organization, and lower than either American military families and Norms on Independence, Achievement Orientation, Moral-Religious Emphasis, and Control. Thus, the findings of this study are fairly congruent with the 1982 study by McGee and Silva of a group of Dunedin NZ families who were compared with American Norms. These earlier investigators (McGee & Silva, as reported in Smith, 1988, p.216) found more Cohesiveness and Expressiveness and less Conflict in NZ families, compared to American families. They concluded that NZ family relationships were more open and supportive and less angry and aggressive. American families, they stated, were more encouraging of independence, more competitive and achievement-oriented. Also, they indicated their results showed that NZ families "held ethical and religious beliefs just as strongly as American families." Interestingly, in this comparisons made for this study, American military families scored higher on Moral-Religious Emphasis (MRE) than either NZ military or NZ civilian families. American military families also scored higher than the Norm for the MRE Subscale. Therefore from these comparisons, it would seem that although NZ families generally (military and civilian combined) score comparable to the Norms on the MRE, neither group holds as strong a moral/religious orientation as American military families. Otherwise, the FES profiles for NZ military and U.S. military families are quite similar.

References

Hunter, E.J. (1982). Families under the flag: A review of military family literature. New York: Praeger.

Hunter, E.J. (1989). Upfront Downunder: An outsider's inside view of New Zealand military families. Wellington, New Zealand: Ministry of Defence.

Moos, R. & Moos, B. (1986). Family Environment Scale: Manual, Second Edition. Palo Alto, CA: Consulting Psychologists.

Nice, S.D., McDonald, B., & McMillian, T. (1981). The families of U.S. Navy prisoners of war from Vietnam five years after reunion. Journal of Marriage and the Family, 5, 431-437.

Smith, A. (1988). Understanding children's development, 2nd Edition. Wellington, NZ: Allen & Unwin.

Interaction Between Drill Sergeants' Families and Units[1]

Pearl Katz, Ph.D.
Walter Reed Army Institute of Research
The Johns Hopkins University School of Medicine
University of Maryland School of Medicine
Uniformed Services University of the Health Sciences

Abstract

An ethnographic natural history of thirty-two drill sergeants and their families throughout their two-year tour has examined the interactive processes which exist among the unit, the family and the drill sergeant. The degree of stress on the drill sergeant and the family is a function of the stability of the family and the social climate in the unit. Family relations affect the quality of drill sergeants' performance either positively through support, or negatively through distraction. The social climate of units mediates the effects of the demands of drill sergeant duties upon the family, and in turn, mediates the readiness of the family to support the drill sergeant in his or her duties.

This research describes the interactive participation of drill sergeants in their families and in their units. The Department of Military Psychiatry of the Walter Reed Army Institute of Research, under the auspices of the DCSPER, has been investigating ways in which aspects of military life facilitate or impede military families' adjustments (Rosen and Moghadam, 1988; Moghadam, 1989; Kirkland and Katz, 1988). Concurrently, research programs have been underway to identify leadership behavior which enhances cohesion and improves mission performance (cf. Marlowe, 1985; Marlowe, et. al., 1989; Kirkland, et. al., 1987). In the course of research into adaptation by the families of drill sergeants it became apparent that soldiers, their units, and their families do not function in isolation from each other. This report describes soldier-family-unit interactions that affect the soldier's morale, the ability of the family to cope, and unit performance.

[1] The views of the author do not purport to reflect the position of the Department of the Army or the Department of Defense.

Drill sergeant duty is believed by many researchers to be highly stressful (Vernon, 1986; Fullerton, 1984). Many drill sergeants expect stress before they begin their tour, have heard anecdotal evidence that marriages frequently disintegrate as a result of drill sergeant duty, and many believe they are stressed throughout much of their tour (Katz and Kirkland, 1989). Although Vernon documented some of the sources of drill sergeant stress in the units and referred to the deleterious effects of that stress on the family (1986), no research has systematically investigated the effects of drill sergeant duty on marriage and family life, nor the effects of marriage and family life upon drill sergeant duty. Nor, importantly, has research systematically examined the positive effects of so-called stressful duty, such as drill sergeant duty upon the family or the positive effects of the family upon the soldier's stressful duty.

Investigation on the causes and implications of stressful military duties and their interactions with the family has suffered from the social organization of social research. Studies on stress of military personnel have usually restricted analyses either to a) psychological effects upon individual soldiers, b) leadership or morale factors impinging upon units, or c) family factors which degrade performance. Each of these contributions to understanding the effects of stress upon performance and readiness has confined its understanding of stress to the causal factors under investigation -- psychological, small-group or family Each contribution has its own body of literature that is minimally shared. Whereas many studies may give lip service to the idea that one factor may affect the other -- such as the unit may impinge upon the vulnerability of the individual soldier, only a few have focused upon the interactive effects of family and duty (Kirkland and Katz, 1989; Moghadam, 1989; Orthner and Pittman, 1989).

Method

Subjects

The data are based upon an ethnographic natural history of thirty-two drill sergeants' families throughout their two-year tour on an US Army Basic Training Post. The primary subjects of the study comprised 58 informants -- 32 drill sergeants (including two drill sergeant couples) and 26 (non-drill sergeant) spouses. Eight of the thirty-two were female, two of whom were single parents. Twenty-seven were drill sergeants for Initial Entry Training; three were drill sergeants for Advanced Individual Training. The demography of the group of drill sergeants reflected the age, gender, ethnicity and length of service distribution of the post's drill sergeant population.

Procedure

The methodologies used in this study were participant-observation and unstructured open-ended interviews. The methodology consisted, first, of recruiting from Drill Sergeants School thirty drill sergeants with families as reseasrch volunteers. Participant observation took place in Drill Sergeant School, on the drill sergeants' training units, in the drill sergeants' homes, and in daily life activities of the spouses. The unstructured open-ended interviews took place with each of the fifty-eight drill sergeants and spouses who were the focus of the study, and with a small sample of drill sergeant instructors at the Drill Sergeant School, the Command Sergeant Major of the Drill Sergeant School, the Command Sergeant Major of the post, 13 NCOs and Officers, Chaplains, and Officers in the Judge Advocate General's command on the post.

The research included a series of five to six-hour interviews with the soldier and the spouse. Approximately four-fifths of these interviews were individual interviews; one-fifth took place with couples. The interviews were spaced throughout the drill sergeants' two-year tours, with the greatest frequency at the beginning. The data base consists of 955 hours with drill sergeants and spouses in twenty-two training companies.

Results

Of the thirty drill sergeant families, twenty adapted to the demands of drill sergeant duty. Seventeen of those twenty families were strong before drill sergeant duty, and remained strong during the tour. Three of those families were not strong before the drill sergeant tours, but improved during the tour. Ten families did not adapt successfully to the demands of drill sergeant duty. Seven of those ten families were strong before drill sergeant duty, but became weaker or deteriorated during the tour. Three of those families were not strong before drill sergeant duty and became weaker or deteriorated during the tour. In four of the families (three of whom were from the weakened families) the drill sergeants either left the service or were seriously planning to leave the service as a consequence of the tour of duty as a drill sergeant; these four had an average of 12.6 years of service.

Criteria for determining the strength of marriages was based upon explicit assessments by the drill sergeants and spouses, and their supportive statements about past and ongoing communication, trust, sexual relations and arguments. Spouses in strong marriages who had satisfactory social support (whether friends, Church, mother or husband) offered significant support for the drill sergeant in his/her unit. Both the drill sergeants and their spouses were acutely aware of the ways in which drill sergeant duty impinged upon the marriage and the family, but they differed in their responses to that impingement.

The units subjected the drill sergeants to different cultural environments and their experiences and perceptions of the environments in their units affected their families. The soldiers' transmission of their attitudes about their units influenced the families' attitudes and perceptions about the units The families' attitudes and perceptions about the units, as well as their internal dynamics and communication, in turn, affected the drill sergeants' concentration, dedication, and performance on the job, and their attitudes toward trainees.

A general pattern emerged from the ways drill sergeants and their spouses perceived the soldiers' companies. Comments by drill sergeants and spouses in the families which adapted were more likely to described the soldiers' units as supportive. Comments about units by families which deteriorated during the drill sergeants' tours were more likely to describe the soldiers' units as indifferent toward the soldiers and their families. Often the drill sergeants' and spouses' perceptions of their units changed over time. These changes proved most often to be associated with changes in the key leaders -- commander, first sergeant or senior drill sergeant.

The contribution of the unit to the interactive process depended to the degree to which the leaders (commander, executive officer, first sergeant, and senior drill sergeant) integrated the three primary facets of the life of the company -- the training recruits, support activities, intra-cadre interactions, (roster duties and details, company administration, taskings by higher headquarters), and soldier-family activities.

In well integrated units the leaders recognized that all the facets are essential to the efficiency of the company, and kept all of them in mind when assigning priorities, communicating with the drill sergeants, and managing time and other resources. In a typical well-integrated company, the leaders encouraged the drill sergeants to call their families when there were breaks in the schedule, reduced manning to a minimum for low-priority activities so that high priority training would have maximum participation, and scheduled activities well in advance so that drill sergeants and their families could predict when they could be together and make plans. In poorly integrated units leaders insisted on full manning for all activities, frequently called off-duty drill sergeants back to the company for trivial reasons, and continually changed schedules and requirements to that families never knew when their drill sergeants would be home.

Discussion

Drill sergeant duty is emotionally demanding and physically fatiguing. Time with families is limited. The degree of stress on the drill sergeant and the family is a function of the stability of the family and the social climate in the unit. An interactive process exists between the unit, the drill sergeant and the family. The extensive time and emotional demands of drill sergeant duty affect both the amount of time with the family, and the quality of active, emotional participation with the family. Family relations affect the quality of drill sergeants' performance either positively through support, or negatively through distraction.

The social climate of units mediates the effects of the demands of drill sergeant duties upon the family, and in turn, mediates the readiness of the family to support the drill sergeant in his duties. Further research will define specific characteristics of the families and the supports for the spouses that enhance the soldiers' participation in the unit.

References

Fullerton, T. D. (1984) Comparing well-being and stress of several high risk Army groups. In Proceedings: Psychology in the Department of Defense. Ninth Symposium, April 1984, 143-152.

Katz, P. and F. R. Kirkland (1989) Command Climate and Drill Sergeants' Families (Washington, D.C: Department of Military Psychiatry, Walter Reed Army Institute of Research).

Kirkland, F. R. and Katz, P. (1989) Combat readiness and the Army family. Military Review, 64, No 4. 63-74.

Kirkland, F. R., Furukawa, T. P., Teitelbaum, J. M., Ingraham, L. H. and Caine, B. T. (1987) Unit Manning Systems Field Evaluation, Technical Report Number 5. Washington DC: Walter Reed Army Institute of Research. September 1987.

Marlowe, D. H. (1985) New Manning System Frield Evaluation, Technical Report Number 1 Washington DC: Walter Reed Army Institute of Research. November 1985.

Marlowe, D. H., Martin, J. and Schneider, R. J (1989) Today's drill sergeant: Dimensions of leadership and training. Army, 39:50-54.

Moghadam, L. Z. (1989) Reciprocal Nature of Work and Family: Perception of the Work/Family Interface and Its Impact on Army Reenlistment Behavior. Unpublished Ph.D. Dissertation Department of Sociology. University of Maryland.

Orthner, D. K. and Pittman, J. F. (1988) Predictors of spousal support for the work commitments of husbands." Journal of Marriage and the Family 50:335-348

Rosen, L. Moghadam, L. Z. (1988) Social support, family separation and well being among military wives. Behavioral Medicine 14:64-70.

Vernon, D. (1986) The Impact of Social Support and Coping on the Relationship Between Work Satisfaction and Well-Being Among Army Non-Commissioned Officers. Unpublished Ph.D. Dissertation. University of Pittsburgh.

The Development and Uses of

The Organizational Climate Survey

Captain Holly K. Burgess, ACSW
Air Force Military Personnel Center

Abstract

The Organizational Climate Survey (OCS) is a management tool which can assist in identifying issues which may affect the productivity, satisfaction, and motivation of personnel. The survey assesses the perceptions of unit personnel on 23 key factors and provides a personnel perspective of how things are going in the unit, i.e., it surfaces both positive aspects and areas which may require additional management emphasis. Although it was developed solely for use by commanders, the OCS has great flexibility and its use has been extended into other Air Force programs. Most recently the OCS was modified for use as a measure for two Air Force programs - the Palace Compete project and Total Quality Management.

The Organizational Climate Survey (OCS) was developed to fill a need expressed by commanders for a method of acquiring information on how their organizations were functioning. The development process involved hundreds of people and all ranks. First, commanders were asked what information about their organizations they thought was important for them to know. The lists of items were consolidated and merged into 23 factors (Table 1). Then an initial bank of items that possibly relate to each factor was developed using several resources and brainstorming sessions. Over 100 people at all management levels independently sorted the item pool and grouped items under the factor they thought the items measured. From these groupings an initial survey instrument was developed and pretested at Air Force bases in San Antonio. After performing a factor analysis, items that were found not to load highly on a factor were discarded. Finally, the survey was administered to over 6,000 individuals (military and civilian) across the Air Force to establish norms for each factor.

The OCS results provide the commander with a systematic assessment of the perceptions personnel have concerning their unit. The argument is sometimes made that attitude/opinion questionnaires are not practical management tools since the data they provide are people's perceptions and not necessarily statements of fact or reality. While this ambiguous relationship between perceptions and reality does at times exist, it does not negate the fact that what an individual perceives is "reality" to that individual, i.e., the individual acts on his perceptions and organizational behavior is influenced by those actions. Thus, the perceptions expressed through the attitudes and opinions in the OCS provide valuable insight into perceived organizational reality. Whether these perceptions are confirmed by a reality check, they are important and need to be addressed if the organization is to remain (or become) healthy.

Table 1

Factors Measured by the Organizational Climate Survey

Sections 1 and 2	
Pay and benefits/economic security	Commitment
Promotion opportunity	Communication
Concern for individual	Identification
Assignment locality	Supervision
Confidence in management	Recognition
Working conditions	Independence
Contribution/participation	Achievement
Group cohesion/worker relations	Interest
Organizational effectiveness	Responsibility
Personal growth and development	Utilization
Sections 3 and 4	
Index of unit development	Job satisfaction
Quality of work life	

Method

Subjects

The OCS can be administered at any organizational unit from a squadron, division, or directorate to an entire wing. The survey is used with both military and civilian personnel. Depending upon the size of the unit and the way the data is to be analyzed, the unit may choose to do a sample or a census. Since most organizations using the OCS have a relatively small population (less than 4,000) using a census for the subjects is usually preferred.

Apparatus

The OCS survey contains a demographics section, four standardized sections, and a section reserved for organization-specific items. Demographics collected are: rank/grade, race, sex, aeronautical rating, and supervisory status. The user may collect further demographic data using the organization-specific section.

Section 1 contains 93 statements (positive and negative) which assess 20 of the 23 key factors which relate to the quality of various aspects of unit life. The respondent is asked to agree/disagree with each of the statements using a 7-point Likert scale, where "Strongly Disagree" and "Strongly Agree" are the end points and "Neither Agree nor Disagree" is the midpoint.

In Section 2 the emphasis turns to the importance the respondent gives to each of the 20 factors which were rated for quality in Section 1. A 7-point scale is used with "Not At All Important" and "Extremely Important" as the end points and "Moderately Important" as the midpoint.

Section 3 consists of two separate measures. The first provides a trend indicator for the quality of the individual's work life. Using a 7-rung ladder as a scale, respondents rate the quality of their work life between "Best" and "Worst" at three time periods: 1 year ago, the present, and 1 year from now. The second measure in this section uses the same ladder scale and asks the member to rate his unit in relation to other units for the same three time periods.

Section 4 uses the 4-question scale (the Hoppock scale) as an indicator of job satisfaction. Since these four questions have been used previously on Air Force-wide samples, the results can be presented in comparison to Air Force averages.

In the section reserved for unit-specific questions the user may submit up to 10 additional questions for analysis. Most organizations use this section to break out the squadrons, divisions, or duty sections of respondents. They may also ask questions relating to the mission of various elements of the organization, such as perceived customer satisfaction or prescription filling turn around time.

Procedure

The Personnel Survey Branch at the Air Force Military Personnel Center provides a camera-ready copy of the survey to the user along with answer sheets designed to be optically scanned. The using organization then administers the survey, collects the answer sheets and returns them to the Personnel Survey Branch for scanning and standardized report output.

Results

Survey results consist of approximately 250 pages of computer printout. Although this may appear overwhelming, the package is structured to facilitate interpretation and make data readily available to the user. The user is also provided the OCS User's Guide (Silva, Hamilton, & Germadnik, 1978) to facilitate understanding and presenting the data. The results are divided into eight reports. In every report except Report 2, the data are reported by the following demographic breakouts: total, officers, enlisted, civilians, supervision, race/ethnicity, and sex. Report 2 presents results in terms of unit total, officer total, enlisted total, and civilian total. Any subgroup with less than 5 respondents is not reported.

Report 1 identifies the average importance and quality for each of the twenty factors measured by Sections 1 and 2 of the survey. This report also contains a representation by factor of the percentage of responses falling below 4, the midpoint of the scale. When at least 30% of the responses for a given factor fall below 4, an asterisk appears next to the quality mean for that factor.

Report 2 focuses on key issues using the individual items from section 1 of the survey. It lists the 10 items which have the highest averages (most positive) and the 10 items which have the lowest averages (most negative). These 20 specific items can provide the user with an additional perspective of strong and weak points within the organization.

Report 3 summarizes the results of the job satisfaction scale. Scores are broken down by demographic categories and the Air Force average for each demographic group.

Reports 4 and 5 outline the results of Section 3 of the OCS survey. Report 4 provides a rating on the quality of work life over the three time periods. Report 5 indicates the relative position of the unit from the worst possible to the best possible.

Report 6 provides detailed results for each of the 93 items from the first portion of the survey. For each item and each demographic group the percent answering each of the 7 responses are indicated. The items are arranged by factor in alphabetical order so items that measure one factor can be viewed together.

The last two reports only apply in some instances. Report 7 presents wing norms for organizations that are wing or equivalent level. The norms are presented as percentile units for each factor so an individual unit can compare its personnel with the total wing. Report 8 is reserved for user-originated questions. For each question, the results are presented by demographic group and percentage of response to each alternative answer.

In addition to the average quality scores on each of the first 20 factors presented in Section 1, the reports present Air Force norms for each factor. These norms include separate comparative values for (1) all Air Force personnel, (2) officers only, (3) enlisted personnel only, and (4) civilians only. Thus, a unit's average quality score on any given factor can be compared to the norm score for that factor. This is an additional input to the commander on the unit's relative strength or weakness in each area.

Discussion

The Organizational Climate Survey has been used successfully by commanders at all organizational levels for over 11 years. Most indicate the survey results are a very useful tool in assessing the health of their unit and pinpointing areas of concern. It has been particularly useful to new commanders who want to get a "feel" for their organization. Because each item on the OCS only weighs on a single factor, it has a modular design. This attribute lends a great deal of flexibility to the OCS as it can be customized for use in other projects.

Recently, it has been used in two Air Force projects. The first project is the Palace Compete Program, a budget initiative within the civilian personnel community being tested at 12 installations. Palace Compete gives local supervisors of civilian personnel control over their civilian funds; the supervisors then determine how to fill vacant positions. The project officers identified five factors (confidence in management, group cohesion/worker relations, organizational effectiveness, promotion opportunities, and supervision) that measure the types of organizational behavior Palace Compete is proposed to influence and added 10 questions specific to the Palace Compete concept. Thus, a standardized Air Force instrument was created which, when used at each test base, will generate Air Force as well as individual base data. When the test period expires in 2 years, the survey will be readministered with required posttest modifications to the questions.

Similarly, the Personnel Survey Branch has recently used the same method to develop an instrument to measure organizational aspects relating to Total Quality Management (TQM). The factors used are communication, confidence in management, contribution/participation, group cohesion/worker relations, organizational effectiveness, and supervision. This survey contains 43 items and also allows the user to add up to 10 additional items. The instrument is available to any unit that wants to survey its personnel in conjunction with implementing a TQM program. It has been tailored to address only TQM-type factors, thereby saving unit personnel much time and effort in building a survey from scratch. Moreover, it is based on a previously validated instrument and provides standardized results while still giving the unit flexibility to ask questions.

The OCS has proven to be an outstanding management tool because it is adaptable to any organizational level and provides a valid picture of the organizational climate. Its best selling points are that it provides direct feedback to the commander at any time, it is solely the commander's decision on the use of the information, and it can be customized for special uses. Users of the instrument enthusiastically say the OCS is clearly a worthwhile and useful survey tool in assessing organizational health. For further information about the OCS, contact HQ AFMPC/DPMYOS, Randolph AFB, Texas, 78150.

References

Silva, W. T., Hamilton, C., & Germadnik, G. (1978). Organizational climate survey user's guide. Unpublished manuscript, Air Force Military Personnel Center, Personnel Survey Branch, Randolph Air Force Base, Texas.

A Proposed Scientific Approach to Building Trust in Our Air Force Leadership

Russell N. Cassel, Colonel, A.F. Retired
CASSEL PSYCH CENTER
Heart Disease Risk Factor Clinic,
1362 Santa Cruz Court,
Chula Vista, California 92010

Abstract

The purpose of study is to provide a scientific approach to the development of trust among Air Force leaders. The critical incident technique is suggested for obtaining objective empirical data pertinent to the building of trust among Air Force leaders, and the further use of Q-methodology to determine the degree of importance of data gathered. The data is proposed for use in training Air Force personnel at varying levels of organization, and assessing the degree of knowledge possessed in relation to building of trust by Air Force personnel.

Excellence in Air Force leadership must be grounded in the effective building of trust in relation to both followers and superiors. It is not something that will occur by chance, and must be pursued with the same scientific rigor that is being used in the development of our arsenal of weapons. It is, of course, presumed that no such valid and empirically derived knowledge presently exists, and that such a knowledge base is of first order of importance to the building of trust. The present proposal involves a four stage process (Stephenson, 1953; and Cassel, 1973):

STAGE I -- the use of critical incident technique to ascertain the precise dynamics involved in building of trust at the various command levels,

STAGE II -- the use of Q-methodology to determine the relative importance of those dynamics, and to forge an assessment instrument to evaluate pertinent knowledge of leaders,

STAGE III -- the development and use of training manuals where factors important to the development of trust are clearly described, and then made part of Air Force training programs, and

STAGE IV -- the use of Q-methodology to assess degree to which Air Force personnel possess the knowledge relative to valid building of trust.

Critical Incident Technique

The "critical Incident Technique" somewhat revised is suggested for the periodic investigation of the structure of trust in relation to the various levels of Air Force organizations (Cassel 1955, and 1958). It is not something that can be guessed about; nor found within the pages of the Holy Bible, or some other mighty treatise, regardless of the level and power status of the writers or authors. By use of this technique criteria in terms of human behavior that is directly related to the building of trust may be effectively identified, and carefully described.

History of Critical Incident Technique

It has a long history that may be traced to the work of Sir Francis Galton, almost a century ago. More recently it was the outgrowth of studies in the Aviation Psychology Program of the United States Army Air Forces of World War II (Flanagan, 1954). One of the first studies carried out was the analysis of the specific reasons for failure in learning to fly that were reported for 1,000 pilot candidates eliminated from flight training schools in the summer and early fall of 1941. A second study, which emphasized the importance of factual reports on performance made by competent observers carried out in the winter of 1943-44 in the 8th, 9th, 12th, and 15th Air Forces. It was concerned with the reasons for failure of bombing missions. Then in the summer of 1944 a series of studies was planned on the problem of combat leadership in the United States Army Air Forces, and which, to be sure, involved the element of "building trust".

A Leadership Q-Sort Test developed by the author is based on 3,667 critical characteristics provided by 400 United States Air Force Pre-flight Cadets by Lt. Joseph H. Hastings for a master's degree at Our Lady of the Lake College, San Antonio, Texas in 1957 (Cassel, 1958). These characteristics are intimately related to the concept of "building trust", and were distributed in the following six clusters in order of importance as follows:

1. Personal Integrity.
2. Consideration for Others.
3. Personal Adjustment and Mental Health.
4. Job Qualifications and Technical Knowledge.
5. Decision Making Skills.
6. Communication Skills.
7. Positive Attitude.

Ascertaining Nature of Trust Building

A series of studies is proposed that parallels the present Air Force education and training programs at the various levels of leadership involvement. An important part of the education and training process would be the gathering of critical incidents related to "building trust" by Air Force leaders at the Non-Com School, Squadron Officers Course, and the Air War College.

This is proposed as an extemporaneous exercise, as opposed to a "take-home" paper assignment, and which would entail no less than a two hour session. The session would be further broken down into three separate sections, with times allotted as follows:

1. One hour for gathering of positive critical incidents,
2. Forty-five minutes for negative critical incidents, and
3. Fifteen minutes for the depicting desirable qualities.

Positive Critical Incidents

This would be broken down into two equally spaced or timed sections for considerations as follows:

First, "Think of the most outstanding Air Force leader at the Squadron Level (as one example, for Squadron Officer's Course) you have known in terms of "building trust". Do not mention his/her name, but carefully list the specific behavior exhibited by that outstanding officer that caused you to think it involved "building trust". Specifically, what did he/she do that caused you to rate him/her as being so outstanding in building trust. Take your time and list the specific behavior or incidents of behavior one after another. Do not be concerned about relating the behavior being listed to the building of trust. Simply list the behavior that caused you to feel the person was outstanding in building trust.

Second (after maybe 30 minutes), now we must continue with the second part of this session. Please stop where you are and begin with this section. Now think of a time when a squadron officer (for that particular level of investigation) has done something you felt should be encouraged because it seemed to be in your opinion involved in building trust for Air Force leadership. Please list these incidents of behavior one after another, and without concern to relating the behavior specifically to the building of trust. Try to be specific in describing the behavior.

Negative Critical Incidents

After maybe a 10 minute break, announce that we are now ready to consider negative critical incidents related to the building of trust. We will spend approximately 45 minutes on this session.

First, think of the most outstanding incident of a squadron officer that in your opinion represented a failure to build trust. Do not mention his/her name, but list the specific incidents of behavior that caused you to feel that this officer was such a failure. Do not try to build a connection between the behavior and the building of trust.

Second (after maybe 20 minutes), now we have arrived at the second portion of this study. This time, think of a time when a squadron officer did something that you thought was not up to par for the building of trust. List the specific behavior as you observed it. Do not try to build a link to the building of trust.

Qualities Related to Building of Trust

Finally, we have arrived at the last portion of this study, and which will entail no more than 15 minutes of time. Here you are to list the qualities you believed are best possessed by a squadron officer for the building of trust. List the qualities one after another. This will conclude the exercise for today, but you may remain as long as you like to complete the assignment.

Cluster Analysis of Critical Incidents

First and foremost it is necessary that all critical incidents of behavior be translated into positive incidents. For all incidents that are stated negatively, they must be changed or converted into positive ones. Here emphasis is placed on the building of trust, not on a failure to build trust. However, it is essential that one learn from failures; so we have identified the negative factors involved, and have converted them to positive forces for training purposes. After the translation into positive incidents is completed, we are ready for the cluster analysis.

Cluster Analysis

Here the incidents are arranged in related groupings, and a count of frequency of occurrence of the incidents is made to determine the prevalence of the incident among the observers. The number of clusters must be reduced to 6; so items must be sorted in terms of that level of generality. A functional assessment instrument should include no more than what would require an hour or class period of time for administering. This probably means 60 items, and to include representative sampling for each of the 6 areas of cluster formations into which the behavior descriptions were assigned (Cattell, 1957; and Wittenborn, 1961).

Using Q-Methodology

Now we have a list of 60 items of behavior that carefully selected persons have identified as being critical to the developing of trust in their opinions. They are distributed equally among the six clusters (part scores on test).

Standardizing the Clusters

Here the same population that contributed the incidents are asked to rank the incidents now translated into positive forces on an 11 point scale, and with numbers as indicated:

11 and 1 -- one item each,
10 and 2 -- three items each,
9 and 3 -- five items each,
8 and 4 -- seven items each,
7 and 5 -- nine items each, and
6 -- ten items.

Interpreting Q-Sort Test on Building Trust

First, it is suggested that the typical Q-Sort Score be used to determine agreement between test results for a particular individual and the norm group of carefully selected successful officers who have taken the test. Since the score involved is a Pearson r (correlation) it represents degree of agreement, and which ranges from full agreement of 1.00, through 0.00 where there is little or no agreement, to -1.00 where there is an opposition to agreement.

Second, in order to depict degree to which the individual displays strengths and weaknesses in relation to the part scores, the test results may be scored by use of R-methodology, and a profile may be made comparing the individual's scores with the Mean for the norm group (this for part scores only)

Manual of Principles for Building Trust

From this data comes a manual that serves as a supplement to leadership development education and training program. In the manual the basic principles deemed as excellence for building trust are enumerated and described in detail (Cassel and Heichberger, 1975).

References

1. Cassel, R.N. (1955). The Air Force Chaplain's Test. Lackland AFB, Texas: Air Force Chaplain's School.
2. Cassel, R.N. (1958). The Leadership Q-Sort Test. Chicago: Psychometric Affiliates.
3. Cassel, R.N.(1973). The Psychology of Decision Making. North Quincy, Massachusetts: The Christopher Publishing House.
4. Cassel, R.N., and Heichberger, R.L. (1975). Leadership Development:Theory and Practice. North Quincy, Massachusetts: The Christopher Publishing House.
5. Cattell, R.B. (1957). Personality and Motivation Structure and Measurement. Yonkers-on-Hudson, New York: World Book Company.
6. Flanagan, J.C. (1954). The critical incident technique. Psychological Bulletin. Vol. 51, No.4, 327-358.
7. Stephenson, W. (1953). The Study of Behavior. Q-Technique and Its Methodology. Chicago: The University of Chicago Press.
8. Wittenborn, J.R. (1961). Contributions and current status of Q Methodology. Psychological Bulletin. Vol. 58, No. 2, 132-142.

Assessment of Stress and Coping Among Basic Military Training Instructors: The Usefulness of the Occupational Stress Inventory

Edna R. Fiedler, Ph.D.
Edwin L. Garwell, Ph.D.
Wilford Hall Medical Center
Curtis Samson
St Mary's University

Abstract

The purpose of this paper is to evaluate the usefulness of the Occupational Stress Inventory in assessing the impact of a chronically stressful job, Basic Military Training, on the individual training instructor.

To most effectively use available personnel, the United States Air Force (USAF) strives to accurately assess an individual's perceptions and reactions to the stress of difficult and/or high risk jobs. Within the civilian world, several instruments are used for high risk jobs. Popular tests include the MMPI or MMPI-2 for pathological indicators, the CPI for general personality predispositions, and job specific behavioral simulations such as in-basket exercises.

Regardless of initial screening, some jobs are important and stressful enough to merit regular on-the-job evaluations of an individual's ongoing stress indicators. Within the USAF, air crews and training instructors are examples of those who receive such regular evaluations. Measures of on-the-job stress such as the MMPI and various stress checklists need to be validated against the specific military population being studied as one cannot automatically assume generalization from a heterogenous normative sample to a pre-selected sample of professionals.

This study looks at one instrument, the Occupational Stress Inventory (OSI) (Osipow, & Spokane, 1987), to see how effective it is at measuring on-the-job stress in a military setting. The OSI was developed as an integrated method of measuring occupational stress, strain and coping resources. Occupational stress is measured by a set of six scales called the Occupational Roles Questionnaire (ORQ). ORQ scales are Role Overload (RO), Role Insufficiency (RI), Role Ambiguity (RA), Role Boundary (RB), Responsibility (R), and Physical Environment (PE). The Personal Strain Questionnaire (PSQ) measures psychological strain and is comprised of four scales: Vocational Strain (VS), Psychological Strain (PSY), Interpersonal Strain (IS), and Physical Strain (PHS). Coping resources are assessed by the four scales that make up the Personal Resources Questionnaire (PRQ). The PRQ scales are: Recreation (RE), Self-care (SC), Social Support (SS) and Rational/Cognitive Coping (RC).

The authors express their appreciation to Tom Swanson for his help in statistical analysis. The ideas presented in this paper do not necessarily reflect the official policy of Wilford Hall or the Department of Defense.

The OSI is based on a model posed by Osipow and Spokane (1983) in which occupational stress, strain, and coping are related and interact. This model postulates that when perceived occupational stressors are equal for two people, differences in coping resources serve to moderate the resulting strain. Thus high stress alone does not necessarily predict strain and strain varies as a function of coping resources. While this model has not been tested thoroughly, preliminary data shows that correlations between stress and strain are slightly but significantly positive and correlations between strain and coping tend to be negative (Osipow, & Spokane, 1983).

More recent research using the OSI found a trend toward lower vocational psychological, physical and interpersonal strain and higher recreational, self-care and rational cognitive coping resources with increasing age (Osipow, Doty, & Spokane, 1985). To date studies have found no main effect for sex (Forney, 1982; Osipow, Doty, & Spokane, 1985).

Questions addressed in this paper include: The adequacy of current norms to adequately represent high risk/high stress samples such as a sample of Basic Military Training Instructors; the relationships between age, IQ and gender to test scores; and, a comparison of scores of TIs recommended for mental health follow-up to scores of TIs not recommended for follow-up.

Method

Subjects

Subjects were 115 TIs (90 male, 25 female) ranging in age from 24 to 50 years old (mean age = 31) with an average IQ score of 114 and an average of 11 years in service. Seventy-three percent of the subjects were married, 12 percent single and 17 percent separated or divorced. A group of eight TIs, recommended for mental health follow-up, were evaluated separately (7 male, 1 female).

Instrumentation and Procedure

Archival data from subjects most recent OSI, MMPI and Shipley were used. The Shipley was used to develop WAIS equivalent IQ scores. The Physical Environment Scale of the OSI was not given to subjects because of earlier information that showed all TI's were homogeneous in this scale. Data was collected on all TIs during required annual mental health evaluations between March and September 1989. The doctoral level credentialed mental health service providers doing the evaluations did not know that the data would be analyzed based upon recommended disposition. Results were analyzed using one sample T-tests and regression analysis.

Results

Male Technical Instructors reported significantly less stress and strain and significantly higher coping scores ($p<.05$) than the normative sample. For males, the means of only one scale of the occupational stress measures, RI, and one coping scale, SC, were not significantly different than the established norms. Table 1 shows the means and standard deviations of male military scores as well as contrasting these means to the normative civilian raw scores equivalent to a T score of 50.

Female Technical Instructors showed the same trend as the males, with seven stress scores and two coping scales being significantly different compared to the civilian female norms. Table 1 shows the female military scores and contrasts them to the normative group.

A correlation between the sample's F-K scores from their MMPI's to their self perceptions of stress was calculated. For every scale, there was a significant correlations ($p < .05$), with increasingly negative F-K scores correlating with lowered perceptions of stress and strain, and higher perceptions of coping resources.

Analysis of the relationship of age, sex, and IQ to OSI scores showed no significant sex or IQ differences on any of the subscales. Significant but modest correlations were found between increasing age and reduced stress as measured by scores on VS ($r = .202$; $p < .05$) and IS ($r = .192$; $p < .05$), as well as increased coping resources as measured by SC ($r = .260$; $p < .05$).

Although there were only eight TI's recommended for mental health follow-up treatment, their scores were strikingly different than those not recommended for mental health follow-up, as can be seen in Table 1.

Table 1
Means, Standard Deviations for
Technical Instructors and the Normative Sample

Scales	MEANS MTI	MN	FTI	FN	MHFU	STANDARD DEVIATIONS MTI	MN[a]	FTI	FN[a]	MHFU
N	90		25		8					
ORQ:										
RO	20*	29	20*	26	23	5	7	7	8	7
RI	22	23	22*	26	26	8	7	9	10	8
RA	16*	21	16*	20	20	5	6	6	6	8
RB	17*	20	19	22	22	6	6	7	3	9
R	25*	27	25	24	26	6	7	7	7	5
PSQ:										
VS	14*	18	14*	19	18	5	5	5	5	9
PSY	15*	20	16*	21	20	6	7	8	8	9
IS	16*	21	18*	22	18	5	6	7	6	6
PHS	13*	17	14*	21	16	4	6	9	7	5
PRQ:										
RE	33*	28	33*	28	27	8	6	7	7	8
SC	29*	27	30	28	24	7	6	7	7	9
SS	43*	41	42	41	37	6	7	7	8	10
RC	41*	37	40*	37	34	6	6	6	6	8

Note. MTI = male training instructor; MN = male normative sample; FTI = female training instructor; FN = female normative sample; MHFU = training instructors recommended for mental health follow-up.
[a] standard deviation for the normative group were computed from Osipow's T scale profiles. * = $p < .05$

When one compares the eight TI's scores to the T scores of the civilian norms, the TI's scores are below or equal to a T score of 50 on six of the nine stress scales measured. Only in this group of eight did we find average coping scores below the T score of 50. Table 1 gives a visual picture of these results.

Discussion

Based on the current study, it is concluded that current available norms do not adequately represent the typical profile or range of scores found in the military sample. To appropriately use such measures as the OSI in the military, norms from the specific sample need to be developed. Otherwise, comparison of an individual's profile to existing norms could easily produce an unrealistic picture of how well the person is doing with occupational stress, personal strain and adequate personal resources.

The effects of social desirability also need to be studied and controlled. In the case of basic military training instructors, their highly negative F-K scores indicate a well-defended group who minimize problems. The significant correlations of their F-K scores to their low OSI stress and high coping scores are interpreted as an indication that they minimized strain and maximized coping in responding to the questionnaire. These tendencies to minimize strain and maximize coping may be adaptive in that the TIs perhaps perceive stress and strain as manageable challenges. Nonetheless, these types of perceptual predispositions need to be taken into account for accurate assessment to occur.

Osipow postulates differences in individual coping resources will moderate differences in perceived strain. The data in Table 1 indicates that TIs recommended for mental health follow up (MHFU) perceive themselves to have many fewer personal resources. The MHFU groups, although not scoring above the normative sample's T score of 50 on six of the nine measured stress/strain scale, do fall below the T score of 50 on all coping resources. Compared to the means of other TIs, the differences between MHFU TIs and other TIs are even more dramatic. In this aspect, Osipow's general model is supported although serious questions about the generalizability and applicability of his normative group remain.

References

Forney, D.S. (1982). Sex and age and the incidence of reported stress, strain and burnout among career development professionals. Unpublished master's thesis, University of Maryland, College Park.

Osipow, S.H., Doty, R.E. & Spokane, A.R. (1985). Occupational stress, strain, and coping across the life span. Journal of Vocational Behavior 27, 98-108.

Osipow S.H. & Spokane, A.R. (1983). A manual for measures of occupational stress, strain, and coping. Columbus, OH: Marathon Consulting & Press.

Osipow, S.H. & Spokane, A.R. (1987). Occupational Stress Inventory. Odessa, FL: PAR, Inc.

Cohesion in Army Reserve Units

A. David Mangelsdorff, Ph.D., M.P.H.
U.S. Army Health Services Command
Fort Sam Houston, Texas

James W. Stokes, M.D.
Academy of Health Sciences
Fort Sam Houston, Texas

Herbert Jacobs, Diplom-Psych.
Federal Republic of Germany

Cohesion and unit morale have been investigated in Army combat units for many years. Relatively little work has focused on Army Reserve units. A measure of cohesion developed by Siebold and Kelly (1988) was modified for Army Reserve units. Psychometric properties of the Reserve cohesion instrument are presented.

Army Research Branch studies of unit morale during the 1940s (Stouffer et al, 1949) were part of an extensive examination of the attitudes of American soldiers during World War II. Measures of group cohesiveness in combat units before the Normandy invasion as well as measures of combat effectiveness of the units during and after the invasion were examined; units with higher cohesiveness performed better. Shils and Janowitz (1948) documented the cohesiveness in the German Wehrmacht in World War II. The more cohesive the unit, the better the unit was able to perform its combat mission. Gal (1984, 1986a, 1986b) has examined unit morale in Israeli combat troops. Ingraham and Manning (1981) have examined the importance of cohesion in the survival of military units in combat. Gal and Manning (1985) conducted cross cultural comparisons of morale in Israeli and in United States Army forces. Mangelsdorff and King (1985) examined multinational efforts at measuring cohesion and motivation in the armed forces. The interrelationships between morale, combat readiness, motivation, performance of military missions, and cohesion suggest the need for further examination of the variables pertaining to a unit's effectiveness.

Scientists from the Walter Reed Army Institute of Research (WRAIR) have been involved in the Headquarters, Department of the Army field evaluation of the Unit Manning System since 1981. The New Manning System (now the Unit Manning System) was developed in 1981 along guidelines derived from the 1979 HQDA ODCSPER-sponsored Army Cohesion and Stability Task Force. The development of the Unit Manning System and its COHORT (Cohesion, Operational Readiness, and Training) unit movement system was undertaken to create military units possessing unit cohesion that would resist psychological breakdown due to combat stress. The evaluations conducted by WRAIR, using survey and interview data at different points in time, confirmed that COHORT units resulted in improved unit morale and enhanced performance (Marlowe, 1985, 1986). The WRAIR research activities involved surveys of "soldier will" (human dimensions thought to be associated with individual combat readiness and psychological sustainability in combat), spouse surveys, a battalion-rotation family-unit-community study, unit interviews, and unit morale and cohesion surveys (Marlowe, 1985). Griffith (1988) described the unit cohesion and morale measures in the Unit Replacement study conducted by WRAIR.

At the U.S. Army Research Institute, measures of cohesion were developed for Army combat units. A long questionnaire on cohesion (the Combat Platoon Cohesion Questionnaire) was developed, then later revised to a short version which became the Platoon Cohesion Index (Siebold and Kelly, 1988). The Platoon Cohesion Index (PCI) is a relatively short (20 questions), focused measure of platoon cohesion which has been used with active duty Army platoons. The questionnaires were administered to 44 Army platoons (752 soldiers). An individual respondent level factor analysis of the 20 items produced three factors which accounted for 60.1 percent of the common variance. The three factors dealt with three types of bonding: horizontal (relations among peers), vertical (relations between leaders and subordinates), and organizational (relations between unit members and the unit as a whole). Each type of relationship has two aspects: an affective (feeling, emotional, reactive) one and an instrumental (action, task, proactive) one. The PCI was developed as a pilot instrument for use in a unit self assessment program to build and maintain cohesion. The PCI is also capable of being incorporated into larger surveys to measure unit climate or other unit conditions.

The purposes of the present research were to develop a survey instrument which would assess cohesion in Army reserve units, to assess the psychometric properties of the instrument, and to examine its uses with reserve units.

Method

Subjects

Soldiers from eight Army reserve units were tested during field training exercises conducted during the summer of 1989. A total of 481 soldiers participated.

Procedure

As part of the field training exercises conducted during the summer of 1989, several reserve OM (psychiatric detachment) teams administered the Unit Cohesion Index to selected reserve units. The Unit Cohesion Index was administered and collected by the OM teams. Selected units were assessed before and after completion of their training.

The Unit Cohesion Index consists of 20 items using 5-point Likert scales which were modified from the Platoon Cohesion Index (Siebold and Kelly, 1988). For 17 of the items, a 5-point Likert format was employed: "Strongly Agree" (1) to "Strongly Disagree" (5).

Analyses of the Unit Cohesion Index results included factor analysis of the 20 items, reliability determinations of the subscales, and analysis of variance comparisons for units which were administered the Unit Cohesion Index on more than one occasion. Evaluations of units by expert observers were compared with the Unit Cohesion Index measures.

Results

Factor Analysis

Responses from the 481 soldiers were submitted to a principal components factor analysis of the 20 items. Four factors with eigenvalues greater than 1.0 were obtained, accounting for 58.3% of the cumulative variance. A Varimax rotation with Kaiser normalization was performed on the factors. Items having an item-total score of .40 and greater were extracted. Three item clusters with at least five items in each were obtained.

Reliability Estimates

The three item clusters were subjected to reliability estimates using the Kuder Richardson procedure to calculate coefficient alpha. The coefficient alphas for the separate scales were: scale 1 (8 items) was .891; scale 2 (7 items) was .805; and scale 3 (5 items) was .760. The scales were labeled as follows - scale 1 (leaders), scale 2 (new members), and scale 3 (unit cohesion).

Level of Unit Cohesion

Scores for individual units on the three scales were evaluated. The mean scores for all units indicated perceived satisfaction with the units and moderately high degrees of unit cohesion. Table 1 depicts the findings. Units that had high levels of unit cohesion were also the units recognized by the expert observers as performing quite well during the training exercises.

- -

Table 1

Reserve Unit Mean Scores on Unit Cohesion Index Scales

Scale	Content	Number of Items	Range of Unit Mean Scores	Unit Mean Score of All Units
1	Leaders	8	11.5 - 22.0	17.0
2	New members	7	9.9 - 14.7	13.2
3	Unit Cohesion	5	7.3 - 10.7	9.9

Note: Low scores are positive: (1) "Strongly Agree" to (5) "Strongly Disagree"

- -

Analysis of Variance Comparisons

Comparisons were made on the scales for units which had taken the Unit Cohesion Index before and after the training. Of the four units on which comparisons were made, there were significant differences for two units on scale 3 (Unit Cohesion). For one unit, there was a significant difference for scale 1 (Leaders), with the unit being more dissatisfied after the training exercises. This same unit showed a significant drop in scale 3 (Unit Cohesion) as well. For the units that had lower scores (greater unit cohesion) before training, even more unit cohesion was reported after the training exercises.

Discussion

Military history has demonstrated the importance of small unit cohesion in the ability of armies to fight and win. Morale and cohesion surveys are routinely administered to determine if an Israeli unit is deployable. Comparisons between the United States and Israeli Defence Force units (Gal and Manning, 1985) suggested that unit morale among United States troops was more closely associated with technical aspects of the unit (condition of weapons and perceived combat readiness), while among the IDF units morale was associated with more human aspects (cohesion, relationships, and confidence in commanders).

As the United States Army involves its Reserve and National Guard units more in Total Army efforts, it becomes important to assess the morale and cohesion of Reserve and Guard units. Mangelsdorff (1988) has examined unit climate and morale in some state guard units; the levels of morale were comparable to those of active duty units.

In the present study, a survey instrument which had acceptable psychometric properties was developed. The level of cohesion found with the Unit Cohesion Index in the reserve units was at least as high as that obtained in active duty units using the Platoon Cohesion Index. The Unit Cohesion Index was sensitive to detecting high levels of cohesion and positive unit attitudes in some of the units which performed well. Further testing of the Unit Cohesion Index is needed to examine other applications for use with reserve units.

References

Gal, R. (1984). Unit morale: Some observations on its Israeli version. In A.D. Mangelsdorff, J.M. King, & D.E. O'Brien (Eds.), Proceedings, Third Users' Workshop on Combat Stress: Cohesion. Fort Sam Houston, TX: Health Care Studies and Clinical Investigation Activity. (NTIS: AD A149035).

Gal, R. (1986a). Unit morale: From a theoretical puzzle to an empirical illustration - An Israeli example. Journal of Applied Social Psychology, 16, 549-564.

Gal, R. (1986b). A portrait of the Israeli Soldier. Westport, CN: Greenwood Press.

Gal, R., & Manning, F.J. (1985). A cross-national comparison of morale assessment: Israeli Defence Forces and US Army. In A.D. Mangelsdorff, J.M. King, & D.E. O'Brien (Eds.) Proceedings, Fourth Users' Workshop on Combat Stress: Lessons learned in recent operational experiences. Fort Sam Houston, TX: Health Care Studies and Clinical Investigation Activity. (NTIS: AD A160228).

Griffith, J. (1988). Measurement of group cohesion in U.S. Army units. Basic and Applied Social Psychology, 9, 149-71.

Ingraham, L.H., & Manning, F.J. (1981). Cohesion: Who needs it, what is it and how do we get it to them? Military Review, 61, 2-12.

Mangelsdorff, A.D., & King, J.M. (1985). Cohesion and motivation: Multi-national efforts in the armed forces. (Consultation Report #85-001). Fort Sam Houston, TX: Health Care Studies and Clinical Investigation Activity. (NTIS: AD A1159940).

Mangelsdorff, A.D. Texas National Guard: Unit Climate Survey Pilot Project. (Consultation Report #88-004). Fort Sam Houston, TX: Health Care Studies and Clinical Investigation Activity.

Marlowe, D.H. (1985). New Manning System field evaluation (Technical report No. 1). Washington, DC: Walter Reed Army Institute of Research.

Marlowe, D.H. (1986). New Manning System field evaluation (Technical report No. 4). Washington, DC: Walter Reed Army Institute of Research.

Shils, E.A., & Janowitz, M. (1948). Cohesion and disintegration in the Wehrmacht in World War II. Public Opinion Quarterly, 12, 29-42.

Siebold, G.L., & Kelly, D.R. (1988). Development of the Platoon Cohesion Index (PCI). Alexandria, VA: Army Research Institute.

Stouffer, S.A., Lumsdaine, A.A., Lumsdaine, M.H., Williams, R.M., Jr., Smith, M.B., Janis, I.L., Stan, S.A., & Cottrell, L.S. (Eds.). (1949). The American soldier: Combat and its aftermath. Studies in social psychology (Vol. 2). Princeton, NJ: Princeton University Press.

The Use of Engineering Psychological Measures In Military Equipment Field Testing

George Anitole and Ronald L. Johnson
U.S. Army Belvoir Research, Development and Engineering Center
Fort Belvoir, Virginia 22060-5606
Christopher J. Neubert
U.S. Army Materiel Command
Alexandria, Virginia 22333-0001

Abstract

The addition of the Engineering Psychologist to research teams has added new dimensions to the evaluation of military equipment in field testing. Historically, measures such as probability of detection and range of detection have been important evaluators of equipment. The Engineering Psychologist has added such measures as visual cues and false-detection rates; the latter is the subject of this report. This paper explains the false-detection variable, and traces its use in two field tests at Fort Devens, MA using military ground observers.

Such standard test variables as probability of detection and range of detection have been accepted for years by the testing community as measures of camouflage effectiveness. Engineers have used these variables to develop requirement documents that prototype camouflage equipment[1/] must meet before becoming standard, fielded equipment. With the advent of the engineering psychologist, now working in these development areas, additional test variables are now coming into use. This report will describe two studies in which the engineering psychological variable, false-detection rate, has been used to complement the standard test variable, range of detection, in the determination of the camouflage effectiveness of woodland camouflage nets and camouflage-patterned Five-Soldier Crew Tents by ground observers. The false-detection rate is defined as the number of times a target, other than the test target, is detected by an observer. In these studies, such detections are rocks, trees, shadows, etc. The number of these detections is a function of how difficult it is to detect the test target. The more difficult the detection task, the greater the number of false detections. The field tests were conducted by the Belvoir RD&E Center at Fort Devens, Massachusetts, in the summer of 1987.

Method

Test Targets: Woodland Camouflage Nets

Four camouflage nets were evaluated. They are described below:

U.S. Army Standard Woodland Hexagon Camouflage Net--The Army standard net has 116 separate patches of incised camouflage cloth (garnish) per hexagon. The garnish is attached to one side of the base net. The net weighs 0.0504 lb./sq. ft., or 34 lb. per hexagon.

Test Net A, 65-Piece Hexagon Woodland Net--This net is similar to the standard hexagon net, except that the garnish patches are larger and number 65 per hexagon. It weighs 0.0474 lb/sq. ft., or 32 lb. per hexagon.

Test Net B, Lightweight Net--This net is made out of unincised, uncoated garnish base cloth which has been sprayed with a pigment coating to lightly fill up the pores of the material. This net is rectangular, and roughly 39 ft x 40 ft., giving it an area of 1560 sq. ft. It weighs 0.0166 lb./sq. ft., or 26 lb. total.

Test Net C, Lightweight Net--This net is made of a light knitted polyester material with a standard camouflage uniform woodland pattern of four colors. This net is rectangular in shape, roughly 25 ft x 35 ft, and has an area of 875 sq ft. It weighs 0.0183 lb/sq. ft., or 16 lb. total.

Test Targets: Five-Soldier Crew Tents

Five, Five-Soldier Crew Tents were supplied by Natick RD&E Center for this study in the following patterns and colors:

Tent A--Standard size four-color uniform pattern repeated every 27.25 inches.

Tent B--Forest Green

Tent C--Expanded four-color uniform pattern repeated every 36 inches.

Tent D--Expanded four-color uniform pattern repeated every 50 inches.

Tent E--Green 483 (cloth dye equivalent to paint color Green 383 used on military vehicles).

Test Sites

The studies were conducted at the Turner Drop Zone, Ft. Devens, Massachusetts, a large cleared tract of land surrounded by a mixture of coniferous and deciduous forest resembling a central European background. Two test sites were selected. Site #1 was located on the western end of the drop zone, so that the morning sun shone directly upon the test targets. Site #2 was located on the eastern edge of the drop zone, so that the afternoon sun shone directly upon the test targets. An observation path, starting at the opposite end of the drop zone from the test target location, was laid out for each site. Each path followed zig-zag, random-length directions toward its test site, and afforded a continuous line of sight to the test target location. The paths were surveyed and marked at approximately 50-meter intervals using random letter markers.

Test Subjects

The test subjects were enlisted soldiers from Ft. Devens, MA. All personnel had at least 20/30 corrected vision and normal color vision. A total of 153 observed the Five-Soldier Crew Tents, and 120 observed the camouflage nets. A minimum of 30 observers were used at each site for each test target in order to employ parametric statistics in analyzing the data[2]. Each observer was used only once.

Test Procedure

The test procedure was to determine the detection distances of the Five-Soldier Crew Tents and the camouflage nets by searching for them while traveling along the predetermined measured paths. Each ground observer started at the beginning of the observation path, i.e., marker S for Site #1 and marker F for Site #2. The observer rode in the back of an open 5/4-ton truck accompanied by a data collector. The truck traveled down the observation path at a very slow speed, about 3-5 mph. The observer was instructed to look for military targets in all directions, except directly to his rear. When a possible target was detected, the observer informed the data collector, and pointed to the target. The truck was immediately stopped, and the data collector sighted the apparent target. If the sighting was correct, i.e., the Five-Soldier Crew Tent or camouflage net, the data collector recorded the alphabetical marker nearest the truck. If the detection was not correct, the false detection was recorded, and the data collector informed the observer to continue looking. The truck then proceeded down the observation path. This search process was repeated until the correct test target was located. The tent and net tests were run consecutively.

The test targets were rotated individually between the two test sites on a daily basis, until all had been observed by at least 15 observers at each site. Their orientations with respect to the sun were kept constant at both test sites. The Five-Soldier Crew Tents and camouflage nets were positioned so that a full side was facing the direction of observer approach.

Results

Detection Data: Woodland Nets

Table 1 indicates which camouflage nets differed significantly from each other.

Table 1. Duncan's Multiple-Range Test--Range of Detection (Meters)

Subset 1		Subset 2		Subset 3	
Group	Mean	Group	Mean	Group	Mean
Army Standard Net	414.41	Net C		Net B	
Net A	469.97	Lightweight	665.99	Lightweight	886.14

The subsets are significant at $\alpha = 0.05$.

The Duncan's Multiple-Range Test separates a set of significantly different means into subsets of homogeneous means The ranges of detection were the shortest for the Army Standard net and Net A. These net ranges were significantly ($\alpha = 0.05$) shorter than the range of detection for Nets B and C. Net C had a range of detection significantly ($\alpha = 0.05$) shorter than Net B.

False-Detection Data: Woodland Nets

Table 2 indicates which camouflage nets had significant rates of false detections.

Table 2. Duncan's Multiple-Range Test--Rate of False Detections

Subset 1		Subset 2		Subset 3	
Group	Mean	Group	Mean	Group	Mean
Net B Lightweight	1.80	Net C Lightweight	3.07	Army Standard Net	4.63
				Net A	5.10

Net B, lightweight net, had a significantly ($\alpha = 0.05$) lower rate of false detections than Net C, lightweight net. Net C, lightweight net, has a significantly ($\alpha = 0.05$) lower rate of false detection than the Army Standard net and Net A.

Detection Data: Five-Soldier Crew Tents

Table 3 indicates which tents differed significantly from each other.

Table 3. Duncan's Multiple-Range Test--Range of Detection (Meters)

SUBSET 1		SUBSET 2		SUBSET 3	
GROUP	MEAN	GROUP	MEAN	GROUP	MEAN
A	327.54	C	351.17	E	674.88
C	351.17	D	387.12		
D	387.12	B	427.71		

The harmonic mean group size is 30.58. The subsets are significant at $\alpha = 0.05$.

The Duncan's Multiple-Range Test separates a set of significantly different means into subsets of homogeneous means. One of the assumptions is that each random sample is of equal size. Since this was not true, the harmonic mean of the group was used as the group size. As seen above, the range of detection was the shortest for tents A, C, and D, and these tents do not differ significantly from each other ($\alpha = 0.05$). Tent E had the longest mean range of detection, and is significantly ($\alpha = 0.05$) different from the other four tents in this respect.

False-Detection Data: Five-Soldier Crew Tents

Table 4 indicates which tents had significant rates of false detections.

Table 4. Duncan's Multiple-Range Test (Rates of False Detection)

SUBSET 1		SUBSET 2	
GROUP	MEAN	GROUP	MEAN
E	2.50	B	3.53
C	3.38	D	3.87
B	3.53	A	4.87
D	3.87		

Harmonic mean group size is 30.58. The rates of false detection for tent groups E, C, B, and D, and B, D, and A were significantly different ($\alpha = 0.05$).

Discussion

As can be seen from Tables 1 and 3, the use of the standard test variable, range of detection, has been very successful in identifying the camouflage nets and Five-Soldier Crew Tents that are most camouflage-effective, i.e., have the shortest detection range. In this case, camouflage nets Army Standard, and Net A, and tents A, C, and D, were the most effective. Before looking at the results of the engineering psychological test variable for nets, false detections (Table 2), one must remember that in a perfect world, the nets and tents that had the shortest detection ranges would have the greatest rate of false detections, while the test items with the larger detection ranges would have the smaller rates of false detections. A review of the net Tables 1 and 2 will show a perfect match with the above expectation. Both the standard engineering test measurement of detection range, and the engineering psychological test measurement of rate of false detections have yielded very clear, clean data about the camouflage effectiveness of the woodland nets being investigated.

The data concerning the camouflage effectiveness of the Five-Soldier Crew Tents (Tables 3 and 4) are not as clear-cut. The range-of-detection data identified tents A, C, and D as the most camouflage-effective, while the false-detection data identified tents A, B, and D. Both test measurements identified tent A as the most camouflage-effective, with the largest rate of false detections and the shortest detection range. Tent E was identified as among the least effective by both test measurements. Duncan's Multiple-Range Test, using both test measurements, has not been able to clearly decide upon the ranking of tents C and B as to camouflage effectiveness. They have been placed in overlapping groups.

From the above data, it has been seen that the engineering psychological test measurements such as rate of false detections are of value in the determination of camouflage effectiveness. These measurements are neither generally known, nor utilized by most test engineers. The addition of the engineering psychologist to the research team has developed new avenues of research, which will ensure more accurate evaluation of prototype equipment.

Conclusions

This paper discusses the role of the Engineering Psychologist in assisting engineers in the field test of military equipment. In addition to the standard field test measurements of range of detection and probability of detection, the engineering psychologist brings additional test measures, such as rate of false detections. A review of two field tests conducted at Fort Devens, MA, indicated that the psychological variable of false-detection rate was most advantageous in the identification of the most camouflage-effective woodland camouflage nets and camouflage-patterned Five-Soldier Crew Tents.

References

1. U.S. Army Combat Developments Command. (10 May 1971). Approved small development requirements (SDR) for a lightweight camouflage screening system.

2. Natrella, Mary G. (1966). Experimental statistics. National Bureau of Standards Handbook 91. Washington, D.C.: U.S. Department of Commerce.

TRIANGULAR PEGS IN SQUARE HOLES: THE LOOSE FIT IN OBJECT DISPLAYS

Dennis B. Beringer and Nancy Lovelace
New Mexico State University

Abstract

A number of investigators have examined object displays in formats including rectangles, triangles, and higher-order polygons. In a number of cases triangular formats were shown to be inferior to bargraphs and other displays. Although these differences have been attributed to emergent features effects in many cases, a more subtle bias relating to the "oblique effect" (differential perception favoring horizontal and vertical lines as compared to diagonal lines) may be present as well. A study was conducted to examine bias in the completion of partial geometric figures having different orientations relative to an implied Cartesian coordinate grid. An overall bias was found to complete right angles as squares without regard to their orientation. A replication produced differential bias according to orientation with horizontal and vertical lines being used to complete figures. These preferences for horizontal and vertical lines in combination with "oblique effect" data reinforce the notion that rectilinear forms of display are unique in the ease and rapidity with which they may be interpreted. These qualities have implications for the design and use of both integrated and object displays.

In the world of multivariate displays and integral dimensions a number of display formats have been explored. Many of these studies have been attempts to integrate multiple dimensions within the context of an object or configural display. Wickens (personal communication) has suggested that the rectangular format of display may be unique in the realm of object displays for several reasons. First, the combination of width and height produces the emergent and integral feature of area. Area is a completely integral combination of height and width; without one of the two contributing variables, the resultant (area) ceases to exist. A second notion is that the rectilinear form has some inherent "goodness" (Gestalt) that makes it a better vehicle through which to convey information that one may wish to integrate. This second notion has some significant underpinnings that are tied to the environment and our perceptual and cognitive development. It is thus worth examining some of the display work performed to date to ascertain how the results may have been influenced by some fundamental aspects of our perceptual system.

Considerable time and effort have been invested in comparing apples with oranges in the discussions surrounding compatibility of proximity (Wickens and Andre, 1988), integral and separable displays (Purcell and Coury, 1988; Sanderson et al., 1989), and figural displays (Beringer and Chrisman, 1987). Beneath all of the higher-level cognitive factors lies a very fundamental factor that influences all of our perception regarding our environment; experience with horizontal and vertical lines. A number of studies have shown that responses to diagonal lines in the environment are different from those to vertical and horizontal lines, not only in humans but in other species as well (Appelle, 1972). Some of this is developmental (Frye et al., 1986) and seems largely to disappear with experience/maturation while other effects persist (Attneave and Thomas, 1977). Those effects that persist seem to bias the organization of the world into horizontal and vertical dimensions, particularly in recall of spatial information (Attneave and Thomas, 1977; Beringer and Hancock, 1989). This bias seems to affect our inherent organization of the world and our expectations of what we see.

The extension of this logic is that our expectations and perceptual development, particularly in a world having a preponderance of horizontal and vertical lines even in non-industrialized locales, are attuned to the horizontal and vertical. Consequently, having people use triangular (or similar) object displays where two of the sides are not in conformance with a rectilinear grid may be like trying to put a triangular peg in a square hole; it is a loose fit at best. Buttigieg and Sanderson (1989) explored a facet of this issue when they implemented a "right-triangle" display. This display depicted departures from normal system operation as a departure from vertical of one of the sides adjacent the normally 90-degree angle (this angle could vary). They found performance obtained using this triangle superior, in general, to that obtained with a similar triangle rotated 120 degrees that did not align the reference angle, and thus two of the sides, with the horizontal and vertical. Hence, some triangular pegs may fit the square hole better than others, particularly if they are perceived as half of the square.

In an attempt to clarify the locus of these expectations, two experiments were conducted to measure the completion of ambiguous "partial" geometric figures. The figures occurred in contexts that were hypothesized to influence the form in which the figures were completed.

Experiment 1: Method

Subjects/Materials /Procedure

The subjects were undergraduate psychology students enrolled in an introductory psychology course at New Mexico State University. There were a total of 16 students who participated. A booklet containing 8 orientations of right-angle figures was used. The figures consisted of the legs of a right triangle with the hypotenuse missing. Four orientations of the figure were aligned with a Cartesian coordinate system (the edges of the page) such that one horizontal and one vertical line could be added to form a square. The other four orientations were aligned at 45 degree angles to a Cartesian coordinate system such that the addition of two diagonally oriented lines were necessary to form a square. The eight figures on the first line of Figure 1 labeled *aligned* are examples of these two types of stimuli. Each of the eight figures was seen two times in the small (3" X 3.5") 16 page booklet. Subjects were given the booklet with written instructions to "complete the geometric figure". After reading the instructions, the subjects used pen or pencil to complete the 16 figures in the booklet.

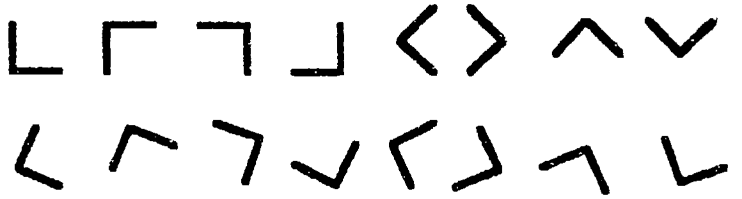


Figure 1. The experimental stimuli.

Results/Discussion

Although an equivalent number of squares and triangles were expected in the completed figures, the overall ratio appeared to favor squares regardless of orientation. The question, however, is two fold. First, do people, on average, favor squares? Yes, there were 150 figures completed as squares as opposed to 76 completed as triangles. There were 31 figures completed in some non-regular manner which were counted as unscorable. The differences were not statistically significant because of individual variability. This means that different people are using different strategies in completing the task, some without regard to the experimental manipulation. Second, are there more "square" people in the sample than those who make predominantly triangles or other figures? Yes, to a barely significant degree. Six responded with more than 87% squares, 2 responded to the experimental manipulation, 1 made triangles, 4 were neutral, and three were unclassified. This small sample size (16) suggested that a replication might be used to increase the data base and sort out these effects.

Experiment 2: Method

Subjects/Materials/Procedure

The subjects were undergraduate psychology students enrolled in an introductory psychology course at New Mexico State University. There were a total of 18 students who participated. The materials were the same as in Experiment 1 with some exceptions. Each of the original figures was presented only once with these eight figures being presented in random order in the first half of the booklet. The second eight figures in the booklet consisted of identical figures which were oriented at angles midway between the orientations of the original angles if these were to be arranged sequentially (see second row in Figure 1). Thus, the second eight figures did not conform to a Cartesian coordinate system, either by being oriented with the H/V axes or by the possibility of aligning the "phantom" side of the triangle with the implied grid. Subjects were each given a booklet which they completed in the same manner as in Experiment I.

Results and Discussion

Data were considered by stimulus condition (favoring squares; favoring triangles) and percentage of squares drawn was used as the dependent variable of a single-factor within-subject design. the percentages were far enough from the extremes that the normal approximation appeared justified and an analysis of variance was used. Overall, subjects produced significantly more squares when stimulus orientation favored squares [mean of 72% versus 42%; $F(1,17) = 9.7$, $p<.01$]. Inasmuch as some unscoreable figures were produced, the percentage of triangles produced was not the complement of the percentage of squares. A separate analysis indicated that subjects produced a significantly larger percentage of triangles when stimulus orientation favored the triangle [mean of 49% versus 14%; $F(1,17) = 11.74$, $p<.005$]. In addition, the mean percentage of squares produced in the nonaligned (favoring neither) condition did not differ significantly from that expected if there was no bias favoring either figure. Thus there is some evidence that use of horizontal and vertical lines to complete these figures is present. Actual frequencies of response are summarized in Table 1.

Table 1. Frequency of figural completion by category for Studies 1 and 2.

		STUDY 1			STUDY 2		
Aligned to favor:	■	▲	TOTAL	■	▲	NEITHER	TOTAL
Figures obtained: ■	74	76	150	52	30	76	158
▲	40	36	76	10	35	59	104
NEITHER	15	16	31	10	7	9	26

Significant subject effects suggested examination of individual strategies. Subjects were subsequently categorized according to their pattern of responses. Those categorized as "responders" had at least 75% of their responses falling in the category predicted by the "use vertical and horizontal lines" hypothesis. A second group consisted of those individuals who unilaterally made squares. Of those remaining, one individual preferred triangles exclusively, two were ambivalent in general, and two could not be classified. Table 2 summarizes the categorization of individuals for both studies. Analysis of the distrubution of classified categorizations across studies indicated a nearly significant departure from a uniform distribution. (Reference χ^2 (3) for .05 = 7.81; obtained χ^2 = 7.42).

Table 2. Categorization of individuals by response strategy.

	Responded to Orientation	Made ■	Made ▲	Neutral or Other	Unclassified
STUDY 1	2	6	1	4	3
STUDY 2	6	6	1	2	3

One interesting note is the way in which "responders" performed in the nonaligned trials; four were ambivalent, one made predominantly squares and one made predominantly triangles. This distribution of categorizations suggests two strategies that are here hypothesized, both favoring Cartesian-coordinate responding. The first strategy involves the completion of figures using horizontal and vertical lines ("responder" bias). The second strategy involves the completion of any right angle as a right quadrilateral ("make squares" strategy). These two strategies appear to account for 2/3 of the sample.

A third sample of equivalent size (18) was very recently collected which exhibited a much more even distribution of responses. The only difference in procedure was that the order of the two major blocks (aligned and nonaligned stimuli) was reversed. Thus it appears that temporal context effects may also influence responding to some degree. The present sample size is too small to allow much generalization, however, and a larger replication designed to sort out some of the effects in question is being conducted.

Conclusions

This series of studies suggests that people have some expectation that a figure containing a right angle will be rectilinear in form, not triangular. The results of the second study suggest that a definite preference exists for the use of horizontal and vertical lines in completing geometric figures. Previous studies have suggested that people acquire spatial information and attempt to force it into rectilinear form either to ease interpretation or to simplify recall (as seen in rectilinear normalization). The implication for spatial displays of information is simple: If we wish to simplify interpretation and recall of information, it may be to our advantage to use the inherent qualities of rectilinear forms to convey the information. While this may limit the number of variables that can appear in a single rectilinear form to two, one may also consider the nesting of rectilinear forms to produce higher-order multivariate displays. It is also clear that individuals invoke strategies to some degree and the factors which determine the selection of a strategy need to be examined. It is as yet unclear to what extent spatial displays may be affected and more data is needed to clarify some of the issues in this area.

Acknowledgement

The authors would like to thank Sharla Fullerton for assistance in stimuli preparation end data reduction.

References

Appelle, S. (1972). Perception and discrimination as a function of stimulus orientation: The "oblique effect" in man and animals. Psychological Bulletin, 78(4), 266-278.

Attneave, F. and Curlee, T. E. (1977). Cartesian organization in the immediate reproduction of spatial patterns. Bulletin of the Psychonomic Society, 10(6), 469-470.

Buttigieg, M. and Sanderson, P. M. (1989). Emergent features and visual display design for two types of failure detection tasks. University of Illinois, Engineering Psychology Research Laboratory, Tech. Report EPRL-89-07 (June).

Beringer, D. B. and Chrisman, S. E. (1987). A comparison of shape/object displays, quasi shape displays, and conventional univariate indicators: Integration effects or the "nearer to thee" effect? Proceedings of the Human Factors Society 31st Annual Meeting, 543-547.

Beringer, D. B. and Hancock, P. H. (1989). Exploring situational awareness: A review and the effects of stress on rectilinear normalization. Proceedings of the Fifth International Symposium on Aviation Psychology, The Ohio State University, 646-651.

Frye, D., Clark, A., Watt, D. and Watkikns, C. (1986). Children's construction of horizontals, verticals, and diagonals: An operational explanation of the "oblique effect". Developmental Psychology, 22(2), 213-217.

Palmer, S. E. (1982). Symmetry, transformation, and the structure of perceptual systems. Chapter 5 in J. Beck (ed.), Organization and Representation in Perception, Lawrence Erlbaum Ass., Inc., 95-142.

Perkins, D. N. (1982). The perceiver as organizer and geometer. Chapter 4 in J. Beck (ed.), Organization and Representation in Perception, Lawrence Erlbaum Ass., Inc., 73-93.

Purcell, J. A. and Coury, B. G. (1988). Asymmetric transfer of training between integral and separable displays. In Proceedings of the 32nd Annual Meeting of the Human Factors Society, 1366-1370.

Sanderson, P. M., Flach, J. M., Buttigieg, M. A. and Casey, E. J. (1989). Object displays do not always support better integrated task performance. Human Factors, 31(2), 183-198.

Wickens, C. D. and Andre, A. D. (1988). Proximity compatibility and the object display. In Proceedings of the Human Factors Society 32nd Annual Meeting, 1335-1339.

Temporal Order of Flashing Lights
as a Navigation Aid

S. M. Luria
Naval Submarine Medical Research Laboratory

Abstract

A set of three lights flashing either simultaneously or sequentially was tested as a navigation beacon. The temporal interval between sequential flashes required for the perception of non-simultaneity averaged about 10 msec. The change in temporal interval required to see a difference increased as the initial temporal interval increased; it averaged about 18 msec at simultaneity and about 26 msec when the initial temporal interval was 33 msec.

The U.S. Coast Guard is considering several novel navigation aids. Among them is a triplet of lights which appear to flash simultaneously when sailing on course. When, however, the ship is off course to the right, the right light appears to flash first, and when the ship is off course to the left, the left light appears to flash first (Brown, 1982). The asynchrony would alert the navigator that he was off course and in which direction; he would then alter course until the lights were flashing simultaneously.

We need to know, however, the minimum temporal interval at which observers reliably recognize that two lights are not simultaneous and can accurately judge which was presented first. Efron (1963) has pointed out that "it is possible for most subjects to be aware that two stimuli are not simultaneous but yet be unable to identify which one is first".

Several studies have attempted to measure the smallest temporal interval needed to perceive temporal order. Most are not applicable; some, for example, used dichoptic viewing (Robinson, 1967) or presented the two stimuli to different visual half-fields (Corwin and Boynton, 1968; Efron, 1963; Hirsch and Sherrick, 1961; Lichtenstein, 1961; Rutschmann, 1966, 1973; Sweet, 1953). Many reported 50% thresholds, which seems inappropriate for the present problem: as a practical matter, we cannot be satisfied with a judgment which is correct only half the time.

Westheimer and McKee (1977) reported the study most relevant to the present problem. They presented two adjacent linear targets to the fovea and found that temporal order could be perceived when one was delayed by as little as 3 msec.

I have determined the temporal interval between three light flashes which most viewers can perceive with reasonable reliability and the magnitude of navigation error signalled by such a system.

Method

Subjects

Ten staff members of the laboratory, ranging in age from 23 to 59, observed. All had 20/25 or better visual acuity with a correction if required.

Apparatus

The obervers viewed three lights flashing every two seconds. At the viewing distance of 6 m, the lights were 0.78 deg apart and subtended 0.1 deg visual angle. The flash duration was about 50 msec. The display was produced by three rotating cylinders (Fig. 1).

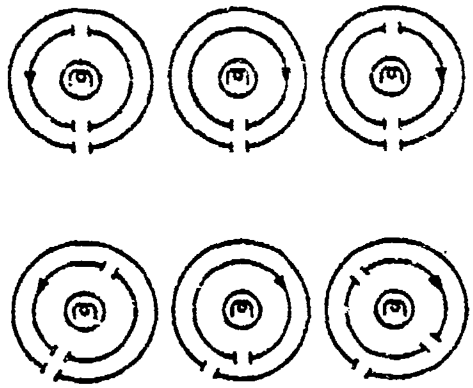


Fig. 1. The interior cylinder on the left rotates counter-clockwise, and the other interior cylinders rotate clockwise. The first and third cylinders rotate at a given speed and produce two flashes for each rotation, while the second cylinder rotates at twice that speed and produces one flash per rotation. This produces simultaneous flashes when viewed head-on and sequential flashes when viewed from off center. The interval between flashes increases as the viewing angle increases.

Procedure

Thresholds were obtained with the method of constant stimuli. A temporal interval was set and the flashing lights exposed until the observer judged whether the lights were simultaneous, or the right or left light appeared first. The lights were occluded while a new interval was set, and so on. The interval at which the observer was correct on every presentation was noted, and, in addition, a probit analysis was carried out to determine the interval perceptible to the viewers 95% of the time.

Results

Table 1 gives the mean temporal intervals between the adjacent flashes required to identify the temporal order.

Table 1
Mean temporal intervals (msec) between adjacent flashes required to identify temporal order on 100% and 95%of the trials

	100%	95%
Mean	9.8	8.4
S.D.	4.4	4.2
Upper Range	16.0	15.0

Difference Thresholds for Various Disparities

At times the navigator will be off course when he first sees the beacons. How much of a change in course is required before he can detect a change in the flashing display?

Procedure

The display was exposed at a viewing angle of either 0, 1, 2, 4, or 6 deg to the right or left of the center line. These corresponded to temporal intervals of 0, 5.6, 11.1, 22.2, and 33.4 msec. The viewing angle remained constant for a random variable period between 5 and 10 sec, after which it was either reduced or increased at a rate of 5 deg/min. The subject was not told when this change began or in which direction it would be changed. He reported when he detected a change and whether the change was toward more or less simultaneity.

Results

Figure 2 shows the mean difference thresholds for each

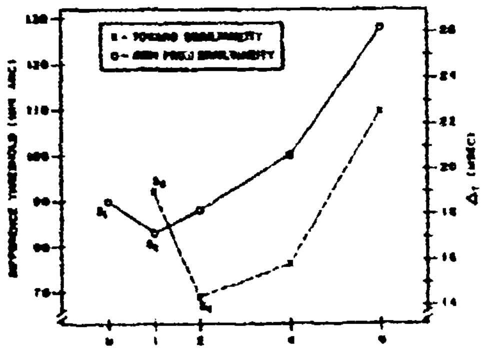


Fig. 2. Difference thresholds in minutes of arc of viewing angle and msec of temporal intergal for different initial degrees of off-center viewing and for changes toward or away from simultaneity. S1, S2, S3, and S4 are referred to in Fig. 3.

starting position both in terms of the change in the temporal interval and the viewing angle for each of the starting positions. It should be noted that these are the means of only nine observers; although all reported that it was a difficult judgment, one could not do it at all at the

4 and 6° conditions. He was not replaced as a subject, because his difficulties constitute an important statement about this type of navigation aid.

As the viewing angle of the starting position from the center line increased (and, therefore, the magnitude of the temporal interval between the flashes increased), it became significantly more difficult to detect a change in the flash pattern according to the Friedman Analysis of Variance by Ranks ($X^2 = 11.2$, $p < .01$). Judgments of decreasing temporal interval appear to be easier than judgments of increased temporal interval, but these differences were significant ($T = 6$, $p < .05$) only for the 4° data, according to the Wilcoxon Matched-Pairs Ranks test.

The curves are not monotonic. There is drop in the thresholds around 1° and 2° after which they rise continuously. The explanation seems quite clear. It is diagrammed in Figure 3. There is a range of temporal intervals around simultaneity which a given subject cannot discriminate. (For most subjects it is 1° to 2°.) When this range is exceeded, he does detect that the flashes are not simultaneous. If the starting position is 1° off center, the resulting temporal intergal is typically too small to detect, but only a small change is required for the subject to respond that the lights are no longer simultaneous. If the starting position is simultaneous, then a larger change is required in order to get out of the range of perceptual simultaneity.

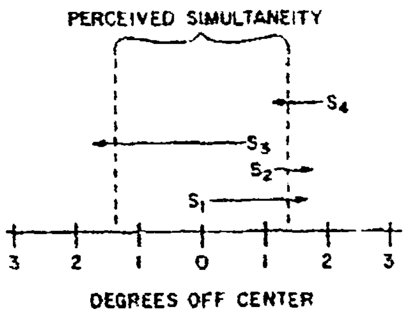


Fig. 3. See text.

If the starting position is 2° off center, this is typically just outside the range of perceptual simultaneity. Thus only a small decrease results in the subjects reporting simultaneity. A much larger change is required if the temporal interval is increasing.

Discussion

These times must be converted to distance measures. For this device, a 12 msec interval corresponds to one degree of eccentricity from the central line of sight. Thus, if an observer requires an 18 msec interval to see that he is off course, that will result in a navigation error of 90 minutes of arc. At a distance of 1000 yards, the obsever will be 26 yards from the center of the channel; at a distance of 5000 yards, he will be 131 yards from the center. Whether this is acceptable of not will depend on the size of the channel to be navigated.

References

Brown, D.M. (1982). Rotating beam single station range design. The Coast Guard Engineer's Digest, 21,43-47.

Corwin, T.R. and Boynton, R.M. (1968). Transitivity of visual judgments of simultaneity. J. Exp. Psychol., 78, 560-568.

Efron, R. (1963). The effect of handedness on the perception of simultaneity and temporal order. Brain, 86, 261-284.

Hirsch, I.J. and Sherrick, C.E., Jr. (1961). Perceived order in different sense modalities. J. Exp. Psychol., 62, 423-432.

Lichtenstein, M. (1961). Phenomenal simultaneity with irregular timing of components of the visual stimulus. Percept. Mot. Skills, 12, 47-60.

Robinson, D.M. (1967). Visual discrimination of temporal order. Science, 156, 1263-1264.

Rutschmann, R. (1966). Perception of temporal order and relative visual latency. Science, 152, 1099-1101.

Rutschmann, R. (1973). Visual perception of temporal order. In Sylvan Kornblum (Ed.), Attention and Performance IV. New York: Academic Press, pp. 687-701.

Sweet, A.L. (1953). Temporal discrimination by the human eye. Am. J. Psychol., 66, 185-198.

Westheimer, G. and McKee, S.P. (1977). Perception of temporal order in adjacent visual stimuli. Vis. Res., 17, 887-892.

Psychophysical Requirements for Three-Dimensional Auditory Displays

Theodore J. Doll
Georgia Institute of Technology
Thomas E. Hanna
Joseph S. Russotti
Naval Submarine Medical Research Laboratory

Abstract

Psychophysical requirements for effective three-dimensional (3-D) auditory displays were studied in a simulated sonar task. Signal-to-noise ratios (SNRs) required to detect a 500 Hz tone in broad-band noise were measured using a loudspeaker array in a free field, as a function of the angular separation of the sources, the correlation of background noises coming from different sources, and listener head movement. The SNR decreased with source separation, and the rate of decrease was significantly greater for uncorrelated noise than for partially or fully correlated noise. The results were essentially identical when the listener's head was restrained versus free to move. Implications for the design of 3-D auditory displays are discussed.

The idea of imparting natural three-dimensional (3-D) qualities to sounds heard over headphones has recently gained attention as a possible addition to man-machine interfaces. Doll, Gerth, Engelman, and Folds (1986) demonstrated that acoustic signals delivered to a listener by headphones can be made to sound like they are coming from a source some distance from the listener. The sound image was made to seem fixed in space, even though the listener was free to move the head. A series of experiments showed that such "virtual" sound sources can be localized with accuracy comparable to real sources stimulating the unaided ears.

Digital synthesizers which impart 3-D qualities to auditory signals are currently being developed, and prototypes have already been demonstrated (e.g. McKinley and Ericson, 1988). Extensive measurements of head-related transfer functions (HRTF's) have been made at the University of Wisconsin in collaboration with NASA (Wightman and Kistler, 1989a). A synthesizer based on these functions has been demonstrated at the NASA Ames Research Center and tested at the University of Wisconsin (Wightman and Kistler, 1989b).

Although digital 3-D sound synthesizers are now available, relatively little is known about their psychophysical requirements. To date, testing has examined only the ability of listeners to localize a single sound source. In most applications, it will be desirable to present multiple sound sources. A potentially serious problem in multiple-source 3-D

This research was supported in part by a U. S. Navy-ASEE Summer Faculty Fellowship held by the first author at the Naval Submarine Medical Research Laboratory. The authors are grateful to Dr. J. V. Tobias for providing resources for the research, and to Ms. Yvonne Masakowski and Mr. Edward Thoma for their support.

auditory displays is that simultaneous inputs from different directions may mask one another. Obviously, a fundamental requirement is that each signal be detectable in the presence of other signals that are presented simultaneously.

The present study examined the conditions under which listeners can detect a signal from one direction when other inputs are simultaneously presented from other directions in a simulated sonar display. The signal was a modulated tone, which is typical of the type of spectral feature used to identify vessels in passive sonar. The other inputs were broadband noise, which simulated sea noise. The signal and noise inputs were presented in a free-field (i.e., a real 3-D auditory display) rather than a synthesized display. The results are therefore not specific to any particular synthesizer, but represent the performance that could be expected from a synthesizer with a high level of psychological fidelity.

The study examined three aspects of 3-D auditory displays: (1) spatial separation of the sources, (2) degree of correlation of the noise coming from the various sources, and (3) listener head movement. There is an obvious design tradeoff involving source separation. Sources spaced too closely will mask one-another, while wide source spacing might unnecessarily sacrifice resolution. It seems likely that the optimal separation of the sources in a 3-D auditory display will be influenced by the characteristics of the signals, such as their correlation, which could be manipulated by signal processing techniques and depends on the application. Head movement was manipulated in order to assess the importance of head coupling.

Method

Seven Coustic model 8822S loudspeakers were arranged in a circular arc of 3.0 m radius centered around the listener's head. The loudspeakers were selected so that their amplitude responses differed by no more than 3 dB from 0 to 20 kHz. The responses were flat to within 3 dB out to 10 kHz, and then rolled off gradually. Testing was conducted in an anechoic chamber.

The signal was a 500 Hz tone with an RMS amplitude of 2 V ac. The tone was amplitude modulated by a 10 Hz sine wave with an RMS amplitude of 3.5 V ac and a 5 V dc offset. The signal was presented exclusively from the center loudspeaker. It had 20 ms rise and fall times which followed a cosine-squared function.

Maskers were presented from either the center loudspeaker alone, or from the center loudspeaker and two others, located at ±10, ±20 or ±40 degrees from the center loudspeaker. The single-speaker condition represented the limiting case of 0 degrees angular separation between loudspeakers.

Outputs of three independent white noise sources were used to produce maskers with specified correlations. The spectrum level of each masker source was 20 dB (re: .0002 dynes/cm^2) at 500 Hz, measured at the origin of the loudspeaker array with the listener absent. With three loudspeakers active, the overall masker spectral levels in the 0.0, 0.8, and 1.0 masker correlation conditions were 24.5, 27.5, and 28.5 dB, respectively. The maskers sounded continuously during testing.

Each subject was tested in 20 experimental conditions. There were 10 Head Restrained and 10 Head Mobile conditions, defined by all possible combinations of three masker correlations (0.0, 0.8, and 1.0) X three loudspeaker angles (±10, ±20 and ±40 degrees), plus one additional condition in which only the center loudspeaker was active.

An adaptive two-alternative forced choice procedure was used to estimate the signal level required for detection in each test block, which included 60 test trials. Detection levels for each condition represent averages over four test blocks. The signal level was decreased a predetermined step size after every three consecutive correct responses, and increased the same step size after each error, which constrained the percentage of correct responses to be 79.4 (d'=1.16).

The subjects were 5 women and 1 man between the ages of 18 and 41 years with normal hearing (< 20 dB HL at octave frequencies from 125 to 8000 Hz, and no more than 10 dB interaural difference at any test frequency).

Results

Figures 1 and 2 show the signal-to-noise ratio (SNR) required for detection in the Head Restrained and Head Mobile conditions, respectively. Three major trends appear in both figures: (1) the SNR decreased with source separation, (2) average SNR over source separation was lower for 0.0 masker correlation than for the 0.8 and 1.0 correlations, and (3) the decrease in SNR with source separation was more abrupt when the masker correlation was 0.0 than when it was 0.8 or 1.0. A comparison of Figures 1 and 2 shows that the results were virtually identical for the Head Restrained and Head Mobile conditions at source separations up to 20 degrees. Beyond 20 degrees, the 0.8 correlation condition showed a more rapid decrease in SNR for the Head Mobile condition than for the Head Restrained case.

A three-factor repeated measures analysis of variance (ANOVA) confirmed the first three observations made above. The main effects of Source Separation and Correlation were both statistically significant ($F(3, 15) = 22.93$, $p < .001$ and $F(2, 10) = 25.51$, $p < .001$, respectively). The interaction of Source Separation X Correlation was also significant ($F(6, 30) = 5.42$, $p < .001$). The main effect of Head Movement and the three-way interaction of Head Movement X Source Separation X Correlation were not significant ($F(1, 5) = 0.02$, $p > .05$ and $F(6, 30) = 0.94$, $p > .05$, respectively). None of the remaining interactions in the ANOVA were significant.

Discussion

The results show that signal detectability in a 3-D auditory display is better with uncorrelated rather than partially or fully correlated noise sources. They also show that increasing the angular separation between sources produces a progressive release from masking, i.e., an decrease in the SNR required for detection. The release from masking with source separation was more rapid when the noise sources were uncorrelated than when they were partially or fully correlated. Head movement had no effect on the SNR required to detect the signal. This result suggests head coupling of the 3-D display is not necessary for the detection task used in this research.

When the sources were 40 degrees apart and the masker correlation was 0.0 the SNR needed for detection in the 3-D

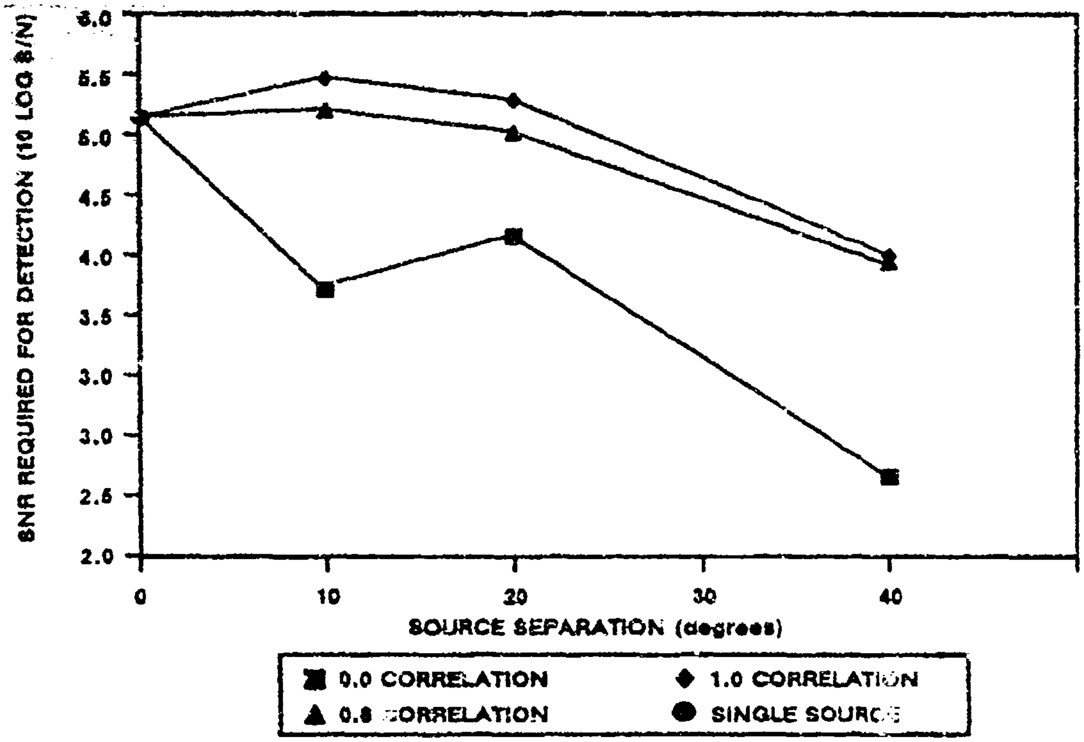


Figure 1. SNR required for detection with the head restrained, as a function of source separation and masker correlation.

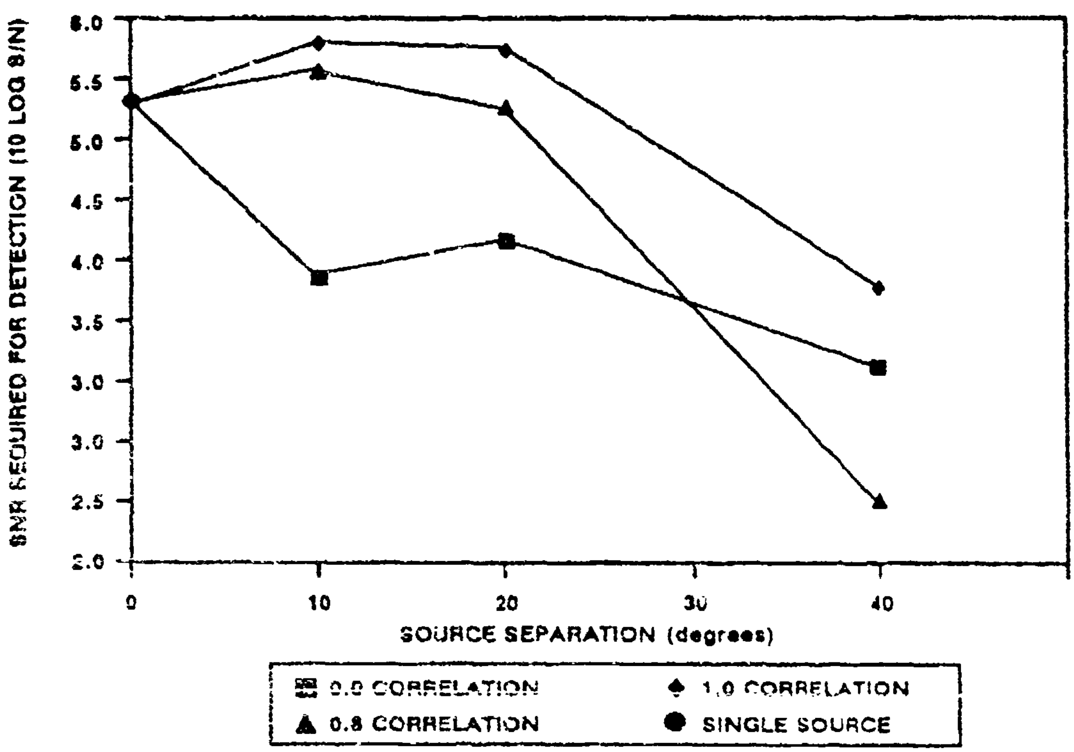


Figure 2. SNR required for detection with the head mobile, as a function of source separation and masker correlation.

display was about 2.5 dB less than that in the single-source condition. However, the absolute signal level needed for detection was still about 2 dB greater than that required when only a single loudspeaker was active. Thus, separating the sources and reducing the masker correlation reduced but did not eliminate masking in the 3-D display. Obviously, if one's only interest were in improving signal detectability in a situation in which signal direction is known, this 3-D display would be inferior to single-source listening (which is similar to auditory displays in submarine sonar systems). However, 3-D displays may have other important applications. For example, they could be used to simultaneously monitor multiple directions for new signals. Doll and Hanna (1989) reported that operators were able to simultaneously monitor an auditory sonar display for any one of four possible signals (i.e., monitor in parallel, like an ideal processor). 3-D displays, which allow the listener to simultaneously monitor multiple directions, might therefore be superior to single-source displays in which the listener must scan to different angular sectors and listen to each in succession. That is, the the mean time to detect a signal from any arbitrary direction might be less for the 3-D display in spite of the 2 dB disadvantage in signal detectability. 3-D displays may also offer advantages over conventional single-source displays for other types of tasks, such as rapid classification and tracking of signals.

The present results show that the sources in a 3-D display must be widely separated (> 40 degrees) in order to eliminate masking. However, the variables examined in the present study represent only a few of the signal parameters that could be manipulated in designing 3-D auditory displays. Further research is needed to examine additional signal parameters to find ways to further reducing masking in 3-D auditory displays, and to determine the effectiveness of these displays for tasks other than signal detection, e.g., discrimination and classification.

References

Doll, T.J., Gerth, J.M., Engelman, W.R., and Folds, D.J. (1986). Development of simulated directional audio for cockpit applications. (Report No. AAMRL-TR-86-014). Wright-Patterson Air Force Base, OH: Armstrong Aerospace Medical Research Laboratory.

Doll, T. J., & Hanna, T. E. (1989). Enhanced detection with bimodal sonar displays. Human Factors, 31.

McKinley, R. L., and Ericson, M. A. (1988, October). A real-time digital auditory localization cue synthisizer. Paper presented at the Symposium on Sound Localization by Humans, National Research Council, Washington, D. C.

Wightman, F. L., and Kistler, D. J. (1989a). Headphone simulation of free-field listening. I: Stimulus synthesis. Journal of the Acoustical Society of America, 85, 858-867.

Wightman, F. L., and Kistler, D. J. (1989b). Headphone simulation of free-fifeld listening II: Psychophysical validation. Journal of the Acoustical Society of America, 85, 868-878.

Developing Effective Coding Schemes for an Advanced Visual Display Technology

Lieutenant Kristen Morton
Harry G. Armstrong Aerospace Medical Research Laboratory

Sarah J. Swierenga, Ph.D.
Logicon Technical Services, Inc.

Abstract

A critical issue in implementing Rapid Communication Display Technology (RAPCOM) is the manner in which information is portrayed or coded within a given display sequence. To evaluate some potential information coding strategies, the current study was conducted in which performance (accuracy and reaction time) on a RAPCOM search/ monitoring task was compared for eight information coding formats at various presentation rates. Results indicated significant differences in both accuracy and reaction time for task performance using the different codes at varying frame durations. This data effectively pares the original set of eight coding concepts down to three--color, font, and spatial orientation--that yielded the best performance on the RAPCOM task. These findings represent a step towards defining efficient coding formats for future applications of RAPCOM technology in operational settings.

An advanced visual display technology has been developed at the Armstrong Aerospace Medical Research Laboratory based on basic research that used multiple sequential frames paradigms for studies of visual search (Chase, 1986) and rapid serial visual presentation (RSVP) reading (Potter, 1984). Rapid Communication Display Technology (RAPCOM) is a new approach (US Patent #4,845,645, 4 Jul 89) for improving information transfer in visual displays. In essence, RAPCOM is a display scheme based on the ability of human observers to effectively process information presented in rapid temporal succession within a single spatial location. (See Figure 1.) As a result, the need for saccadic eye movements is eliminated, yielding a significant savings in time for information transfer that may be critical to users (e.g., fighter pilots) operating at the limits of their sensory capabilities. In fact, Matin and Boff (1988) have reported RAPCOM savings from two to five times over comparable conventional display arrays. In sum, RAPCOM technology offers the designer several potential advantages over conventional display approaches: enhanced speed of information transfer for some types of information processing; reduced space requirements for the display, since several pieces of independent information can be presented by a single RAPCOM window; control over information sampling sequences, which can encourage superior display monitoring strategies; and ease of implementation in operational environments, where displays are already largely computer driven (Matin and Boff, 1988).

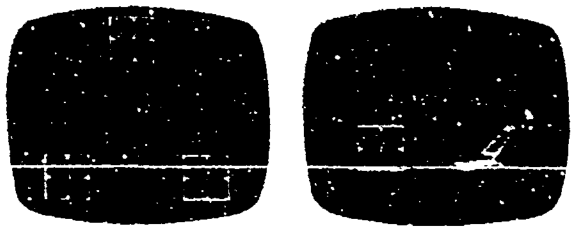


Figure 1. A conventional display with three windows and the corresponding RAPCOM display (not to scale).

While previous research efforts have demonstrated that RAPCOM technology is a viable concept (Matin, Boff, and Donovan, 1987; Matin and Boff, 1988; Osgood, Boff, and Donovan, 1988; Payne, 1988), its feasibility for application in complex system environments remains to be shown. The key to successful application of RAPCOM technology lies in whether or not the information presented can be portrayed so that successive frames can be independently discriminated by the system user. A critical issue here is the manner in which the information in a given frame is portrayed or coded with respect to any other frame of information in a RAPCOM sequence. In other words, if information is not adequately "coded", it will likely appear to the system operator as a meaningless jumble of letters and numbers. A coding strategy is needed that will optimize the usefulness of RAPCOM-presented information and enhance its potential for operational crew system applications.

A review of the coding literature revealed that for normal adults, language as a code or format takes processing priority over all others because of its vast overlearning (Teichner, 1979). Thus, linguistic and contextual cues may satisfactorily act to code some types of information (Chen, 1986). However, when the information presented consists of numeric values or sets of related variables (e.g., in-flight parameters in a cockpit environment), contextual coding alone is unsatisfactory. Several other coding alternatives have been suggested: color of a single item or segment of data, brightness contrast of certain pieces of information, symbols or shape coding, alphanumerics and different types of fonts, size, flashing of an item, inverse video of field and ground, line type, and redundant coding, which may enhance the effects of individual coding schemes (Narborough- Hall, 1985; Christ, 1975). Alone or in combination, several of these coding concepts may be feasible for incorporation into a RAPCOM scheme.

The objective of this study was to compare several potential information coding strategies within a RAPCOM paradigm, and to select three or four that showed the greatest promise for future research. Eight coding conditions were developed to help subjects discriminate frames of information within RAPCOM trials. In the "label" condition, each RAPCOM frame had one of four labels: "A", "B", "C", or "D". In the "frame color" condition, the surrounding area of each frame was colored one of four colors: cyan, pink, yellow, or red. In the "text color" condition, the border of each frame and the text within each frame were colored one of four colors: blue, red, green, or black. In the "font" condition, the text within each frame was displayed in one of four fonts: italic, bold, underline, and shadow. In the "background pattern" condition, the surrounding area of each frame contained one of four patterns: horizontal lines, diagonal lines, dots, and checker board. In the "frame shape" condition, each frame had one of four shapes: triangle, rectangle, diamond, and oval. Finally, in the "spatial offset 1" and "spatial offset 2" conditions, each frame was offset slightly from the center point of the display. The difference between these two conditions was that one displayed a central focal point and the other did not. Previous RAPCOM studies used rectangular frames with black borders and text, and a white surrounding area.

Method

Subjects

Four subjects, two females and two females, ages 22 through 29, participated in this experiment for five sessions. All subjects reported having normal color vision and 20/20 corrected visual acuity, and were right-handed. All subjects had previous experience with the RAPCOM paradigm, and received five training sessions on this particular task.

Apparatus

This study was conducted in the Design Effectiveness Technology (DEfTech) Laboratory at the Harry G. Armstrong Aerospace Medical Research Laboratory, Wright-Patterson AFB, OH. An Apple Macintosh II laboratory computer system with an enhanced color graphics monitor provided overall control of stimulus presentation and data acquisition. Programming was done primarily in HyperCard, with X Command supplements written in C. The response panel was a standard Apple keyboard, placed on a computer table in front of the seated subject. Four response keys were designated with the labels A, B, C, and D. Subjects were permitted to choose which hand and finger would operate the keys. The computer recorded accuracy, reaction time in milliseconds, and also recorded which keys had been pressed. This information and other summary data were printed out on an Apple Laserwriter II immediately following each session, so the experimenter could monitor the subject's performance between sessions.

Procedure

The experimental design consisted of two within-subject factors, coding condition and display speed. In total, subjects ran eight coding conditions at five display speeds. Coding conditions included: label, frame color, text color, background pattern, frame shape, font, and two conditions in which frames were slightly offset spatially. The frame display times used were: 250 ms, 200 ms, 150 ms, 100 ms, and 50 ms. Before each run, subjects had up to five minutes to memorize the code that was in effect for that run. The stimuli consisted of RAPCOM frames containing three-digit numeric values ranging from 000 to 999, with 200-800 considered "in-bounds". Four arbitrary parameters, A, B, C, and D, were defined to correspond to the four frames of information presented in each trial. In an operational setting, these parameters would be meaningful (e.g., air speed, altitude, etc).

Subjects were told that their task was to respond as quickly and as accurately as possible to out-of-bounds values presented during RAPCOM trials. They were instructed to determine which of four values presented per trial was out of the preset boundaries, and to identify which of the four constant parameters that value corresponded to. For example, in the "frame color" condition, a cyan surround corresponded to parameter A, a red surround to parameter B, a pink surround to parameter C, and a yellow surround corresponded to parameter D. Responses were given by first pressing "Enter" to acknowledge that a value was out of bounds, then pressing "A", "B", "C", or "D" to identify which parameter was out of bounds, and finally pressing "Enter" again to signal that the response was completed. Reaction times for all three button presses were recorded.

Subjects completed five one-hour experimental sessions. Sessions consisted of twenty-four runs of ten trials each--a total of 240 trials per session. Within each session, the order of conditions was randomized. Within each trial, the order of parameters presented was randomized. Also within each trial, the three-digit values were randomized between 000 and 999, with one of the four values being out of bounds. All trials were scored. Each run consisted of a three-second ready signal followed by ten cycles (trials) of four RAPCOM frames. The run would not continue until the subject made three key press responses following each trial. Subjects received feedback regarding their reaction times and accuracy following each run.

Results

A repeated measures analysis of variance was conducted to assess the effects of frame duration and coding condition on accuracy (% correct) and reaction time. Reaction time values used for analysis were calculated by adding the reaction times for the first two button presses for each trial, representing the cognitive portion of subjects' responses. For F-tests involving more than one degree of freedom in the numerator, the Greenhouse-Geisser degree of freedom adjustment was used.

The effect of frame duration on reaction time was significant, F (1.5, 4.5) = 39.34, $p < .002$, as was the effect of frame duration on accuracy, F (1.5, 4.5) = 70.30, $p < .001$. As frame duration decreased, reaction time increased and accuracy decreased, indicating that RAPCOM performance was better at slower display speeds. See Figure 2.

The effect of code on reaction time was significant, $F(1.6, 4.7) = 11.97$, $p < .016$, and the effect of code on accuracy was also significant, $F(1.6, 4.7) = 24.61$, $p < .001$. Code did not interact with frame duration for either reaction time or accuracy. Paired comparisons were used to assess differences between specific codes in terms of reaction time and accuracy. For reaction time, the background code was significantly slower than all other codes, $p < .044$. For accuracy, the background, label, and frame shape codes were significantly less accurate than the five other codes, $p < .033$. See Figure 3.

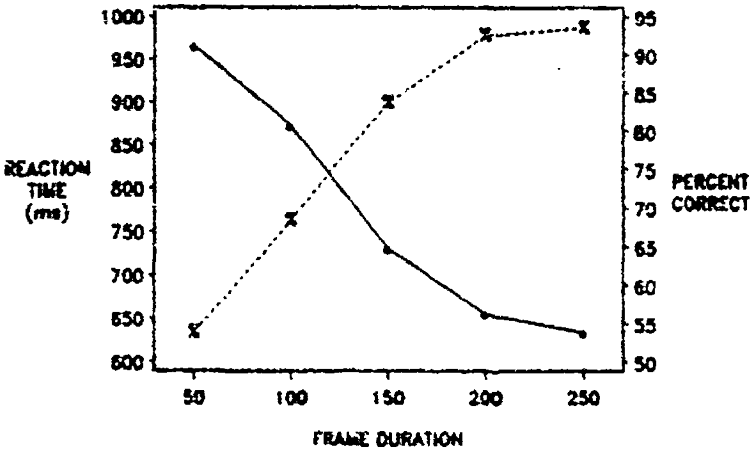


Figure 2. Effect of frame duration on reaction time and accuracy performance measures using RAPCOM displays.

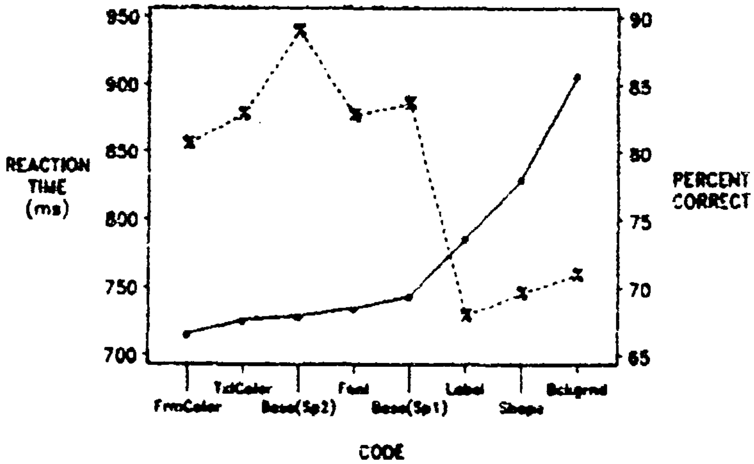


Figure 3. Effect of coding schemes on reaction time and accuracy performance measures using RAPCOM displays.

Discussion

The initial hypothesis was that certain codes would be more intuitive to learn and use, and consequently would be more effective for coding information in RAPCOM displays. Data analysis supports this hypothesis, finding significant differences between codes in terms of performance--accuracy and reaction time. The frame color, text color, font, and two spatial orientation codes proved to be the most effective, while the background, label, and frame shape codes were least effective in facilitating RAPCOM performance. Further, these five effective codes can be collapsed into three general coding strategies: color, spatial orientation, and font.

Further research on coding issues will compare variations of these three most promising coding strategies, which facilitated the fastest reaction times at the fastest display speeds, and yielded the highest accuracy. The focus of the next study will be to further narrow and define the feasible coding approaches, examine the effects of redundant coding, and ultimately select an optimal RAPCOM coding format. Other lines of research planned for this technology include: (a) evaluating the effect of RAPCOM coding schemes in dual-task paradigms, and (b) testing the effects of high-g acceleration on RAPCOM performance in a dynamic environment simulator.

Potential applications for RAPCOM technology include military crew stations, air traffic control centers, nuclear power plants, space stations, and any other complex, computer-controlled systems in which human operators must receive and manage large amounts of information (Matin et al., 1987). However, full realization of this potential is contingent upon defining an efficient, effective RAPCOM coding format. This data effectively pares the original set of eight codes down to three coding strategies--color, font, and spatial orientation--that yielded the best performance on the RAPCOM task. Thus, these findings represent a step towards applying RAPCOM technology in operational settings.

References

Chase, W. (1986). Visual information processing. In K. Boff, L. Kaufman, and J. Thomas (Eds.), Handbook of perception and human performance (pp. 28-1 - 28-71). New York: John Wiley and Sons.

Chen, H. (1986). Effects of reading span and textual coherence on rapid-sequential reading. Memory and Cognition, 14 (3), 202-208.

Christ, R. (1975). Review and analysis of color coding research for visual displays. Human Factors, 17 (6), 542-570.

Matin, E., Boff, K., and Donovan, R. (1987). Raising control/display efficiency with rapid communication display technology. Proceedings of the Human Factors Society 31st Annual Meeting, 258-262.

Matin, E., and Boff, K. (1988). Information transfer rate with serial and simultaneous visual display formats. Human Factors, 30 (2), 171-180.

Narborough-Hall, C. (1985). Recommendations for applying colour coding to air traffic control displays. Displays Technology and Applications, 6 (3), 131-137.

Osgood, S., Boff, K., and Donovan, R. (1988). Rapid communication display technology efficiency in a multi-task environment. Proceedings of the Human Factors Society 32nd Annual Meeting, 1395-1399.

Payne, D. (1988). Performance in a visual monitoring task with serial and simultaneous display formats (Contract No. F49620-87-R-0004). Dayton, OH: AAMRL/HEX.

Potter, M. (1984). Rapid serial visual presentation (RSVP): A method for studying language processing. In D. Kieras and M. Just (Eds.), New methods in reading comprehension research (pp. 91-118). New Jersey: Lawrence Erlbaum Associates, Publishers.

Teichner, W. H. (1979). Color and visual information coding. Proceedings of the Society for Information Display, 20, 3-9.

THE REDUCTION OF CENTRAL-VISUAL OVERLOAD IN THE COCKPIT

Lisa F. Weinstein, Ph.D.
University of Illinois at Urbana-Champaign

Abstract

A time-sharing experiment investigated the use of non-traditional flight displays for the reduction of central-visual overload in the cockpit. The possibility of a compatibility mapping between the type of task and the presentation location of the task in the visual field was also explored. The results indicate that a non-traditional flow-field display allowed for accurate control and the most efficient time-sharing. Limited evidence was found to support a compatibility hypothesis.

The increasing complexity of technology found in the aircraft cockpit has dramatically increased the number of displays from which a pilot must extract information. As a result, the central vision of the pilot has become overloaded which could lead to decision making errors and increase the potential for accidents. The experiment to be described tested two possibilities for reducing central-visual overload. Based on the relative strengths of central and peripheral vision, the first possibility involves off-loading some visual tasks to peripheral vision. It is well established that with retinal eccentricity motion sensitivity is degraded less rapidly than acuity. One hypothesis in the current research was that when central vision is overloaded and visual-spatial tasks are to be off-loaded from central vision, performance on motion-tasks will be degraded less then performance on object-recognition tasks.

Leibowitz (1986), in presenting a dichotomy between focal and ambient vision suggested that a wide field-of-view including peripheral vision is best suited for motion-judgment tasks, and that when presented in this manner motion tasks can be performed with few processing resources. Leibowitz's theory would suggest that peripheral vision would be better at tracking and at time-sharing than central vision.

The second method suggested for reducing central-visual overload involves the use of non-traditional flight displays. The current study examined the use of peripheral vision and flow-field displays. The use of peripheral vision horizon displays (PVHD) has been widely discussed and a variety of attitude indicator displays for aircraft have been designed and tested using this channel. Hameluck (1988) presents a critical review of the literature on the PVHD. If these displays are effective, they would provide an alternative method for presenting attitude information to a pilot while reducing central-visual load.

For the purposes of this research, a display will be considered to be flow-field if it meets three criteria: 1) extension across the central and peripheral visual fields; 2) presentation of information in a way that is pictorially consistent with the information available in the real-world;

3) incorporation of characteristics of displays which have been called ecological (i.e., splay and compression, see Gibson, 1979). The symbology in the flow-field display was designed in part based on Roscoe's (1968) principle of pictorial realism. Roscoe's idea suggests that realistic pictures require less interpretation and processing (or in Gibson's terms, more direct perception) and should therefore require fewer processing resources. The experiment examined the attentional demands, in a dual-task flight simulation, associated with the use of non-traditional displays.

Method

Displays

Three experimental displays were employed which are pictured in Figure 1. The figure illustrates the three displays showing high altitude, and a right bank with the aircraft heading right. The first display is a central display similar to a traditional attitude indicator, except that the current central display was outside/in or world referenced. An horizon line was located in the center of the display and extended the width of the flight envelope. The task was to maintain the symbol upright and on the center of the horizon line.

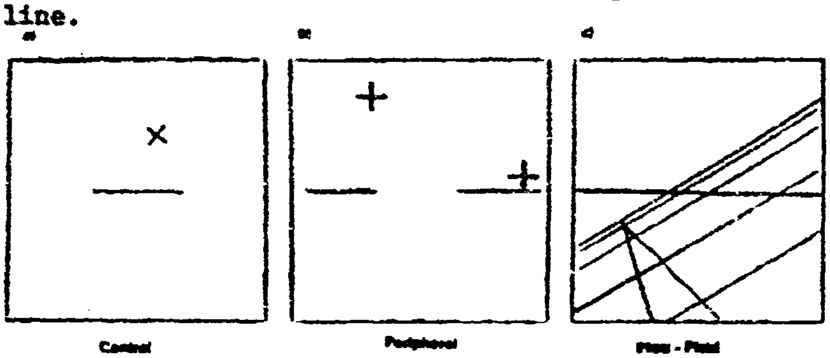


Figure 1: Displays illustrating rolling right, high altitude and heading right of course.

The second display, shown in Figure 1b, was a peripheral display. The display consisted of two crosses located 20 degrees of visual angle apart. Behind each cross (during level flight) was an horizon line that extended the width of the flight envelope. In order to maintain straight and level flight, subjects had to maintain both crosses centered on the horizon lines. The task required that the subjects view both crosses simultaneously which could only be accomplished with peripheral vision.

The third display, shown in Figure 1c, was a flow-field display. The display provided splay and compression information in the form of horizontal lines, and a full-field horizon line. Altitude information was available from the changes in the splay angle, as well as from the distance between the horizontal lines. The task was to maintain straight and level flight at an altitude displayed at the beginning of the trial.

Flight task

The flight task consisted of a two-axis compensatory tracking task. Subjects were asked to maintain straight and level orientation while being buffeted about by pseudo-random wind gusts. During each of the three experimental sessions, subjects completed 10 trials. Trials lasted four minutes, during which subjects hovered as opposed to moving through the world. Subjects controlled altitude through pitch, and heading through roll using first order control dynamics. Root mean square error in tracking was recorded, as well as root mean square control velocity.

Secondary tasks

While flying, subjects periodically performed one of two secondary tasks, an object-recognition task and a motion-discrimination task. In the object-recognition task subjects were asked to identify one of two shapes (i.e., a square, a diamond and a triangle) which appeared either in the center of the display or at both of the peripheries for 500 msec.

The motion-discrimination task was in most ways directly analogous to the object-recognition task. Two rectangular windows filled with random dots appeared either side by side in the center of the display or at the two peripheral locations of the display. The dots in the windows were either moving down, not moving, or moving up on the display. The duration of presentation of the stimuli was 500 msec. Subjects were asked to indicate the direction of motion in one of the two windows. Accuracy and reaction-time to classify the events was recorded.

Apparatus and subjects

The experimental displays and the secondary tasks were presented on a Silicon Graphics 3020 IRIS computer system. A high resolution (1024 x 768 pixels) color monitor presented the displays. Subjects were seated such that the monitor subtended a visual angle of 31 (horizontal) x 24 (vertical) degrees. Flight control was input by using a 2-axis spring-loaded displacement stick. Responses to the secondary tasks were input by a keypad. Thirty right-handed male students from the University of Illinois volunteered as subjects, and ten were assigned to each display condition. All were flight naive, and reported normal or corrected to normal vision.

Results

Flight data

The root mean square (RMS) errors (altitude and heading) and the control velocities (altitude and heading) were recorded during single-task and dual-task flight trials. MANOVAs were conducted on these measures to assess the differences in time-sharing performance with the four secondary tasks and the three display types. In RMSE altitude, there was a small advantage for the central display ($F(2,27)=3.25$, $p=.054$), relative to the flow-field display (see Figure 2). The RMSE heading analysis revealed a significant task type by location interaction ($F(1,27)=5.35$, $p<.05$) (see Figure 3). When the tasks were presented centrally, they effected flight performance equally. However, when presented in peripheral vision, the shape tasks produced significantly greater decrements in flight performance than the motion task.

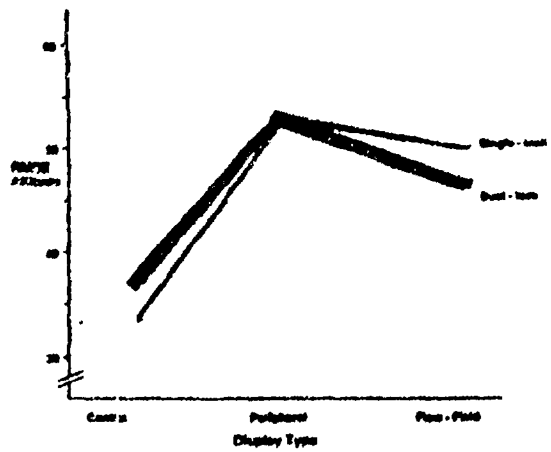


Figure 2: Root Mean Square Error (altitude) as a function of display type.

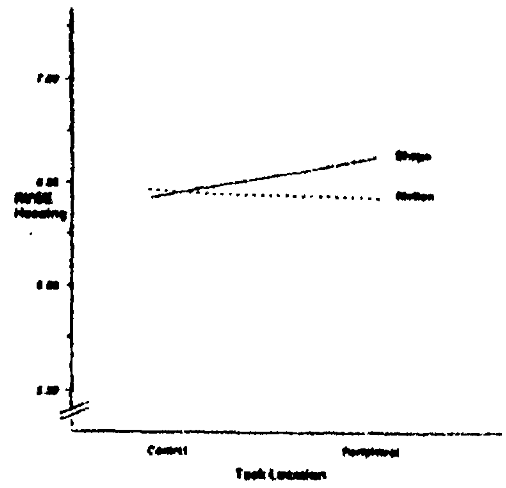


Figure 3: Root Mean Square Error (heading) as a function of task location and task type collapsed across display type.

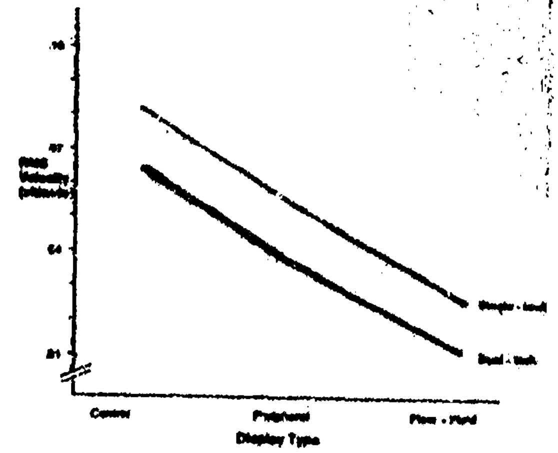


Figure 4. Root Mean Square Velocity (altitude) as a function of display type

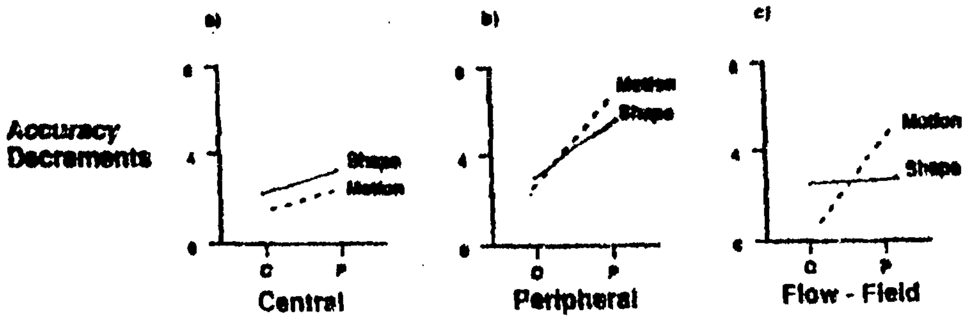


Figure 5: Secondary task, dual - task decrement scores by display type. Figures a, b and c depict accuracy decrements (percent correct). Each figure illustrates the task location by task type interaction.

The RMS control activity (altitude) analysis showed a significant main effect of display type ($F(2,27)=19.52$, $p<.001$). In order to achieve accurate control with the central display large control actions were needed (see Figure 4). Much smaller control actions were needed with the flow-field display.

Secondary-task data

The MANOVAs on the accuracy decrement scores revealed that the location of the task has no effect on accuracy with the central display, but has a significant effect with the peripheral ($F(1,9)=9.41$, $p<.05$) and flow-field displays ($F(1,9)=5.87$, $p<.05$). With the latter two displays, a central-secondary task resulted in smaller accuracy decrements than a peripheral-secondary task. The absence of a difference with the central display suggests less interference between two central tasks than between two peripheral tasks.

Conclusions

Several general conclusions can be drawn from the previous discussion. First, tasks presented in peripheral vision are consistently more difficult, whether they involve shape or motion discrimination. This finding contradicts previous suggestions that peripheral motion processing required few if any resources. Second, central and peripheral vision do not constitute separate resources. If they did, then time-sharing one central and one peripheral task would produce the best performance. The current results showed two central tasks producing the best time-sharing. Third, the RMSE heading data suggest that when off-loaded to peripheral vision the motion task interfered less with tracking than did the shape task.

Finally, although the central display resulted in the best single-task altitude control, the flow-field display was found to be superior to the central and peripheral display in several ways. The flow-field display allowed for fairly accurate control, reduced control activity, and very efficient time-sharing. Future research will further investigate the use of flow-field displays as a possible solution to central-visual overload in the cockpit.

This research was funded by a grant from NASA Rotocraft Human Factors Research Branch (Contract No. NAG 2-308). Sandra Hart was the contract monitor.

References

Gibson, J.J. (1979). An ecological approach to perception. Boston: Houghton Mifflin.

Hameluck, D. (1988). The peripheral vision horizon display: A critical review. Doctoral dissertation, York University, Toronto.

Leibowitz, H. (1986). Recent advances in our understanding of peripheral vision and some implications. In the Proceedings of the 30th Annual Meeting of the Human Factors Society. Santa Monica, CA.

Roscoe, S.N. (1968). Airborne flight displays for flight and navigation. Human Factors, 10, 321-332.

Knowledge Acquisition in Building Expert Systems: Applying the Human Factors Perspective to Achieve a Principle-Driven Process

Captain Jay A. Horn

USAF Academy

Abstract

This paper examines the problem of knowledge acquisition in the development of expert systems. While many of these expert systems have been billed as possessing the capability to robustly cope with the demands of their particular domains, the literature reveals that often these systems are less capable than expected by the expert system designers, often due to problems during knowledge acquisition. This paper proposes a principle-driven process, based on human factors principles, of knowledge acquisition for building expert systems.

The science of artificial intelligence and its related subfields of expert systems, robotics and machine vision, and natural language processing have recently exploded into the forefront of scientific and engineering research and applications. Expert systems, defined as knowledge-based systems that emulate expert performance to solve significant problems in a particular domain of expertise (Sell, 1986), are currently receiving a great deal of attention.

Because of this interest (and increased human factors interest, see Wheeler, 1989), the potential benefits expert systems possess, and the failure of expert systems technology (in general) to live up to this potential, it is appropriate to examine how expert systems are built and how the process can be improved. This paper examines the basic process of how expert systems knowledge acquisition is conducted, with recommended improvements of that design process in light of the human factors perspective.

Problem Statement

Despite the claims of their developers, many expert systems have yet to realize their potential as useful domain experts that provide useable knowledge to system users. For example, Dreyfus and Dreyfus (1986) describe numerous examples of expert systems that did not successfully meet their intended goals. I believe the current process of acquiring domain information ignores some basic design principles that human factors engineers use daily.

Central to this concept is an appreciation for the information requirements of the expert when performing in the domain. Addis

(1985) has indicated that a more complete understanding (rooted in theory and establishment of in-depth techniques) of the human factor in the process of expert system design is required if the science is to move forward.

Kidd and Sharpe (1988) are also concerned with the lack of research investigating human expertise in artificial intelligence-related fields. They regard the current generation of expert systems as experiments, which have focused impressive amounts of computational power on specific problems in highly isolated domains, but have yet to achieve sufficient basis in theory to allow growth into diverging domains. They write:

> Success of the system is directly determined by the appropriate representation and application of specific knowledge from that domain to solve an isolated problem. Despite the vast amount of data now available as a result of these experiments, we are still unable to explain the "why" or "how" of successful systems or to predict for which other domains and tasks the current techniques will work. This is because no theory of tasks or domains currently exists (p. 147).

These represent a sample of the comments researchers have made urging additional research and improvement of the expert systems development process.

The Nature of the Problem

Allen (1987) describes five examples of traditional methodologies of building expert systems. Each of the methods require a step to gather and define relevant domain knowledge that contributres to performance. This identification and structuring of the domain knowledge necessary for adequate performance is gathered using otfen painstaking and inefficient interviews, known as knowledge acquisition (Boose and Gaines, 1988). The entire success or failure of an expert system is usually determined in the knowledge acquisition phase. Improvement of this step would make a significant improvement to the expert system development process.

Expert Systems Knowledge Acquisition

The knowledge acquisition process is typically performed by a knowledge engineer (or engineers) who interviews and questions the domain expert (or experts) using real and hypothetical examples of cases to determine the heuristics used by the experts. When these "rules of thumb" are stated in an "if-then" format, a rule-based model of domain expertise is created (Klahr and Waterman, 1986).

This set of rules forms the basis for the rule-based expert system. If this set of rules is incomplete or in error, the system will not perform as anticipated. Revision of the rule-set and additional expert interviews are required to bring the system up to acceptable levels of performance.

This problem of a long and tedious cycle of interviews and rule-set revisions has been referred to as the "knowledge acquisition bottleneck" (Feigenbaum, 1977, p. 71). Waterman (1986) has observed the knowledge acquisition step can extend "over a period of many months." New methods of knowledge acquisition that allow the expert to communicate directly with the machine to develop rule-sets have been developed to circumvent this problem (see Quinlan, 1987, for some examples). However, the need to refine and correct the knowledge initially gained is still evident in spite of these new methods (Klahr and Waterman, 1986).

Many reasons have been offered to account for the difficulties encountered in knowledge acquisition. For example, experts are believed to have difficulty articulating deeper-level rules, experts may forget to include rules they take for granted (commonsense), and/or experts may make incomplete or erroneous assumptions. Generally, the difficulties focus on the problems inherent in the interview process and getting domain experts to translate largely intuitive processes into verbal representations.

Why Is There a "Bottleneck"

Artificial intelligence and expert systems researchers have failed to notice their traditional approach used in building rule-based expert systems constrains the domain analysis to only those elements supported by the current level of technology. Herein lies the cause of the "bottleneck" problem. This bias towards the limitations of expert systems technology can prevent system designers from adequately understanding the domain and those information elements critical to successful performance in that domain. Figure 1 depicts the steps involved in this traditional approach.

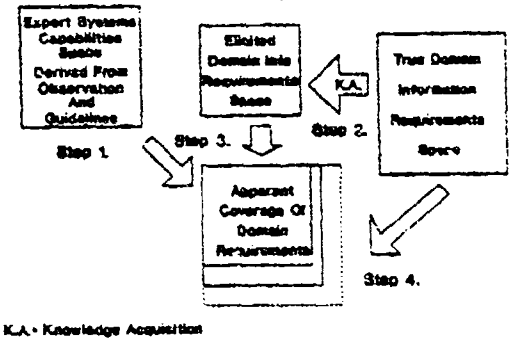


Figure 1. Traditional Approach to Building Expert Systems.

In Step 1, the Expert Systems Capabilities Space (ESCS) is developed in the mind of the knowledge engineer (KE) through experience. This represents what the KE believes expert systems can do and the types of information they can support. The content of the ESCS is limited by the knowledge engineer's contact and understanding of expert systems.

As she performs the knowledge acquisition process in Step 2, an Elicited Domain Information Requirements Space (EIRS) is developed based on rules elicited from the True Domain Information Requirements Space (TIRS) via the domain expert. The EIRS is only as complete as what the domain expert states and is included as important by the KE. It is necessarily an incomplete subset of the TIRS.

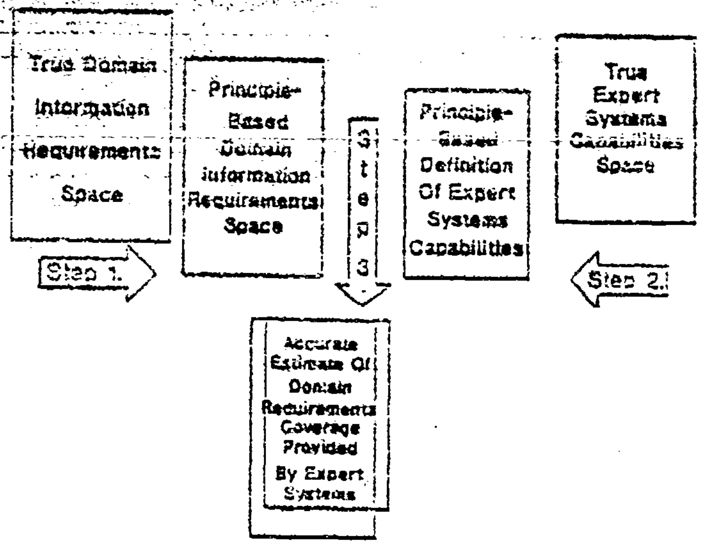


Figure 2. Principle-Based Approach to Building Expert Systems.

In Step 3, the ESCS are mapped onto the EIRS, apparently covering the entire domain. However, in Step 4, after actual use, system performance is often found to be less than expected. If it were possible to map the ESCS onto the TIRS, a significant portion of the domain would be found to be unsupported. This lack of support explains much of the resulting poor system performance.

Conclusion and Recommendations

The knowledge engineer's inherent bias towards the system prevents him from fully understanding and appreciating the domain information requirements. The KE's attention to what the system can support often distracts her from identifying what information is necessary and critical to domain performance. Therefore, the most relevant (but often unsupported) information aspects of a domain are not included in the expert system rule-set, resulting in sub-standard system performance.

An appropriate model of human cognitive task execution (see Horn, Koubek, and Cacioppo, 1989) applied to the knowledge acquisition process would result in a more complete and valid representation of the domain. Figure 2 outlines the use of such a model in the development process. Steps 1 and 2 use theory to specify and define both the domain elements and capabilities of expert systems. What is important to note is that the domain information requirements are defined first, to insure the all relevant and critcal aspects of the domain are represented. Step 2 then casts the capabilities of the expert system inlight of those domain-specific information requirements. Therefore, in Step 3, mapping the two together results in an accurate estimate of the degree of support (coverage) that can be expected by the expert system.

In conclusion, by utilizing the human factors principles of user-centered design and identification of domain information requirements, it is appears possible to mitigate the problem of system-bias in the knowledge acquisition process. Research to investigate this hypothesis is recommended to improve the critical function of knowledge acquisition and result in more capable expert systems.

References

Addis, T. (1985). Designing Knowledge-Based Systems. Englewood Cliffs, New Jersey: Prentice-Hall.

Allen, M. (1986). The Development of an Artificial Intelligence System for Inventory Management Using Multiple Experts. Oakbrook, Illinois: Council of Logistics Management.

Boose, J., and Gaines, B. (1988). (Eds.) Knowledge Acquisition Tools for Expert Systems. London: Acedemic Press.

Dreyfus, H., and Dreyfus, S. (1986). Mind Over Machine. New York: The Free Press.

Feigenbaum, E. (1977). The art of artificial intelligence 1: Themes and case studies of knowledge engineering. STAN-CS-77-621, Department of Computer Science, Stanford University.

Horn, J., Koubek, R., and Cacioppo, A. (1989). A theory-based methodology for analyzing domain suitability for expert systems technology applications. In G. Salvendy and M. Smith (Eds.), Designing and Using Human-Computer Interfaces and Knowledge Based Systems. Amsterdam: Elsevier.

Kidd, A., and Sharpe W. (1988). Goals for expert systems research: Analysis of tasks and domains. In D. Moralee (Ed.), Research and Development in Expert Systems IV. Cambridge: Cambridge University Press, 146-152.

Klahr, P., and Waterman, D. (1986). Expert Systems: Techniques, Tools, and Applications. Reading, Massachusetts: Addison-Wesley Publishing Company, Inc.

Quinlan, J. (1987). Applications of Expert Systems. Reading, Massachusetts: Addison-Wesley Publishing Company, Inc.

Sell, P. (1986). Expert Systems- A Practical Introduction. New York: John Wiley & Sons.

Waterman, D. (1986). A Guide to Expert Systems. Reading, Massachusetts: Addison-Wesley Publishing Company, Inc.

Wheeler, J. (1989). More thoughts on the human factors of expert systems development. HFS Bulletin, 32 (12).

A Preliminary Comparison Between Neural Network and Rule Based Maneuver Selections in an Air Combat Training System

Daniel C. McMahon, M.S.
Timothy E. Goldsmith, Ph.D.
University of New Mexico

Abstract

The purpose of the present research was the comparison of a rule based expert system against a neural network expert system in the domain of air combat maneuvering. Differences between the architectures were explored and hypotheses as to causes of differential performance were made. In addition, a methodology for converting rule based systems into a neural network was also presented. Both models were compared with expert fighter pilots on a transfer task. The neural network agreed with maneuver selections made by expert fighter pilots 2.5 times more often than the rule based system. These findings were explained in terms of the ability of neural nets to generalize maneuver selections to novel airspace conditions. Other implications of these results were also discussed.

The goal of this research is to develop a neural network model that selects offensive and defensive maneuvers for the domain of air combat maneuvering. Specifically, the present research will compare a rule based architecture, Air Combat Expert Simulation (ACES), against a neural network architecture, Tactical Air Combat Intelligent Trainer (TACIT), as models of Air Combat Maneuvering (ACM). Each model is designed to select basic fighter maneuvers (i.e., maneuvers that enhance one's tactical position) during ACM. The ultimate goal of the present work will be the implementation of a stand alone personal computer training system to be used as an adjunct to flight and classroom instruction during the training of undergraduate fighter pilots.

Current work on TACIT is a direct descendent of previous work done by Goldsmith and Schvaneveldt on ACES (Goldsmith & Schvaneveldt, 1987). ACES selected classical maneuvers using a series of production rules defined as condition-action or if-then statements. An example might be, if my opponent is in front of me and our range is less than 1500 feet then fire guns. In all, there are 38 production rules that select 12 offensive and 5 defensive classical maneuvers. During operation, ACES samples an ACM scenario and attempts to find matches between rules and ACM conditions. When a match is found a rule is selected. Because ACM is such a dynamic decision making environment it is quite often the case that conditions for more than one rule are satisfied. When more than one rule has been selected a tie-breaking or conflict resolution procedure is applied in an attempt to discriminate the best maneuver. ACES' conflict resolution scheme is to select the most general rule. That is, the maneuver rule that has the fewest airspace descriptors defined (i.e., fewest number of conditions defined) is selected.

A problem that many conflict resolution schemes have is that they operate at a syntactic level. Another way of stating this is that the information contained within the condition half of tied rules, that was initially used to make selections, is not used in the conflict resolution process. Rather, meta-information such as how many or how few conditions are specified by the set of rules becomes the basis of discriminating between potential actions. It may be the case that sub-groups of airspace conditions within rules carry important discriminatory information beyond that found in measures of the generality or specificity of the individual rules. Many conflict resolution procedures ignore this semantic aspect of a rule. One approach to capturing specific configurations of conditions in airspace scenarios is to write large numbers of unique rules. Unfortunately, in dynamic environments such as ACM this can lead to vast numbers of rules. In an attempt to overcome this limitation of rule based expert systems a neural network model of ACM was implemented.

A major advantage neural network models have over rule based systems is that they are designed so that performance degrades gracefully given incomplete or noisy specifications of input information. Neural networks are designed such that they select an action in terms of previously learned mappings of stimuli to responses. The neural network used here is able to generalize to previously learned responses given incomplete or novel instances of airspace conditions. In other words, to the extent that a new set of conditions describing an ACM scenario is similar to a pattern that has already been trained into the network, a similar pattern of processing will occur across the network resulting in a similar response (Arbib, 1987).

Another advantage neural networks have over rules is that the network is trained in much the same way Undergraduate Pilots (UPs) are; —namely, both UPs and the network are exposed to a large number of ACM scenarios and asked to make the appropriate maneuver selections. Upon selection of maneuvers both UPs and the network are given response feedback until they learn to make appropriate maneuver selections.

Since the maneuver rules in ACES where based on interviews with expert pilots as well as validated against experts it was thought that there was an important knowledge base contained within those rules. Therefore, the decision was made to train a neural network using the information contained within the production rules. In order to train a network using this information it was necessary to re-conceptualize rule based maneuver selection as a classification task. The condition half of a rule was viewed as the prototype or central tendency for making a maneuver classification. That is, the conditions were viewed as the attributes that composed the prototype while the actions were thought of as the categories into which each prototype was classified. Table 1 illustrates the transformation of an acceleration maneuver (AC) into a prototype with its associated maneuver category.

Production Rule	Network Training Pattern
select (AC) if my_aircraft(0/45,_,B,_), opponent_aircraft(_,_,F,5/7), relative(2/12,_,-400/250).	prototype: [22.5,_,0,_,_,_,1,6,7,_,-75] maneuver class: [1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0]

Table 1. A Production Rule Conceptualized as a Prototype

It should be noted that the underlined blank spaces in both the production rule and the prototype correspond to conditions that are unspecified. In cases where there was a range of possible values for a specific airspace condition the average value was used in the network training pattern. The maneuver categories are simply represented as a binary vector where each value corresponds to a maneuver. In this example acceleration is the first value in the vector and is represented as 1 while all other maneuvers are represented by 0s. The 38 rules in ACES were reconfigured as 38 prototypes and associated category names. The 38 prototype-category vectors were used as the input-output patterns on which the neural network was trained.

In a previous study Goldsmith and Schvaneveldt (1987) presented a series of 40 airspace scenarios to expert fighter pilots who were asked to select maneuvers. The airspace scenarios presented to the expert pilots and the maneuver selections they made will be used to validate the production rules and the neural network maneuver selection architecture. The data collected from expert pilots will also be used to compare the performance of both models.

Method

Network Topology

The specific network to be used was a multi-layered perceptron model with a nonlinear activation function. The model was trained using the standard delta-rule or back-propagation gradient descent learning algorithm (Pao, 1989). The network had three layers driven by twelve input conditions. The first hidden layer had 12 nodes, the second hidden layer had 20 nodes, and the output layer had 17 nodes (a 12-17-20 network). All nodes were completely connected to successive layer nodes resulting in a total of 688 interconnections. Each connection is operationalized as a weight which reflects the degree of association between each node and every other successive layer node.

Network Training

The network was trained by presenting each of the maneuver prototypes, applying the non-linear activation function, computing an error term, and modifying the network interconnections. It took 17,500 blocks of maneuver presentations (i.e., 1 block is a single presentation of the entire set of 38 maneuver prototypes) for the network to learn to correctly classify 36 of 38 maneuvers. The reason the network did not learn all 38 maneuver classifications was that there was an insufficient number of connection weights (i.e., the network was too small).

Results

The rule based system agreed with expert pilot maneuver selections in 10 of 40 transfer scenarios (25%). The neural network maneuver selections agreed with expert pilot selections in 27 of 40 transfer scenarios (67%). Using a binomial distribution, the probability of either model agreeing with expert pilot maneuver selections at the levels reported here due to chance is essentially zero ($p < .0001$). While both models performed well, it should be noted that the neural network agreed with pilot maneuver selection over 2.5 times better than the production system.

Discussion

While both models agreed with the maneuver selections made by expert pilots at levels significantly better than chance, the performance of the neural network was much better than the rule based system. Initially, this finding seemed puzzling in that the neural network was trained to recognize the conditions from the production system as maneuver categories. In other words, the neural network contained the same information as is found in the production system. The reason for the difference lies with the ability of neural networks to gracefully degrade given incomplete or noisy input. Neural networks do not have the architectural constraint of having to make explicit condition matches nor do they have to apply conflict resolution for a maneuver classification to be made. Rather, maneuver selection is driven by the degree of similarity between the pattern of activation in the network and the patterns initially trained into the network. A useful metaphor is to think of each transfer condition as having a distance between itself and the 38 prototype patterns initially trained into the network. The shorter the distance, the greater the probability that the transfer pattern will be placed into that category.

The production system, on the other hand, yielded tied productions on the majority of transfer scenarios. The result was that the conflict resolution strategy arbitrated final maneuver selection on most transfer trials. This was supported by the fact that when fighter pilot maneuver selection was compared with the set of all tied productions, agreement was nearly 95%. It appears that the conflict resolution strategy employed by ACES degraded the agreement with first maneuver selection by expert pilots.

It is the case, however, that both fighter pilots and tutoring systems act upon first maneuver selection. This makes the first selection per scenario the critical performance test for a tutoring/expert system. In this regard, the neural network implemented in the TACIT system performed at a superior level to the production system. While it may be the case that an increased number of productions sensitive to unique airspace conditions and/or inclusion of semantically based conflict resolution strategies may enhance a rule based model's ability to generalize to novel ACM scenarios, neural networks already have this ability as a basic component of their architecture.

In conclusion, a method has been demonstrated whereby rules from a production system can be used to train a neural network. The main benefits are reflected in enhanced performance as a result of the model's ability to generalize to new instances of airspace conditions.

References

Arbib, M. A. (1987). Brains, Machines, and Mathematics. New York: Springer-Verlag.

Goldsmith, T. E. & Schvaneveldt, R. W. (1987). Representing and training expertise in air combat maneuvering. (AFHRL TP-86-60).

Pao, Y. H. (1989). Adaptive Pattern Recognition and Neural Networks. New York: Addison-Wesley.

Establishing Instructional Strategies for Advanced Interactive Technologies

Daniel J. Muraida
Jonathan M. Spector
Air Force Human Resources Laboratory
Brian E. Dallman
3400 TCHTW, Lowry AFB

Abstract

Instructional Systems Design (ISD) in the military is based on a behavioral model of learning and a systems approach to engineering management. The current ISD process is adequate for some traditional educational settings. However, advances in cognitive science, new interactive training technologies, and increasingly complex training requirements indicate a need to broaden the learning theory that drives ISD and to adopt a more integrated approach to training development. The Advanced Instructional Design Advisor (AIDA) is a proposed method to test and integrate new instructional strategies for teaching complex tasks using advanced technologies.

Instructional Systems Design (ISD) is the established process for designing and developing instructional materials for military training. This process works well enough for relatively simple materials to be presented in a classroom setting (Montague, 1988). However, the ISD process is bankrupt, if not broken, when it comes to providing specific guidelines for teaching complex tasks in an interactive setting (Muraida & Spector, 1989).

Advances in educational technologies have brought the limitations of ISD to light. Given a variety of available and affordable presention media, how should materials be selected and presented so as to maximize understanding, retention, and transfer? The response to this question has been to give the task to highly qualified and creative individuals and see what they produce (Montague, Wulfeck, & Ellis, 1983). The answer has been that imagination is the limit to instructional design. What is lacking is a systematic method to determine whose imagination is effective, when, and why.

The embarrassment of the riches of educational technology is that there is not a theoretically based justification for using anything other than chalk and an instructor's imagination. What is lacking is a well developed instructional theory.

There is obviously a great deal that we do not know about how people learn, so the best that instructional theorists can hope to get from learning theory is a good direction to pursue. The key, then, to developing an instructional theory will be the testing, refining, and refuting of those instructional prescriptions suggested by learning theories. The Air Force Human Resources Laboratory's Advanced Instructional Design Advisor (AIDA) is intended to provide the key instructional design validation component necessary to developing an instructional theory.

Problem

It is difficult to claim that the instructional design techniques used by Air Force personnel are inappropriate or fail to work. There is sufficient evidence to indicate that motivated subject matter experts (SMEs) can use the existing instructional design model to produce a training system which meets the training need (Montague, 1989). Experienced SMEs invariably choose a robust design mediated by the most flexible medium known: classroom lecture/laboratory and instructor-based. The problem is not that current ISD procedures are inadequate to the task of producing effective classroom setting, instructor-based technical training.

Part of the problem is that there is a scarcity of knowledge and experience in designing instruction to be delivered using new technologies. Even the best professionals produce wasteful or inappropriate designs.

There has been no comprehensive evaluation of the relative cost-effectiveness of training systems produced by different categories of instructional designers. Recently, an established aerospace firm with an excellent track record produced a course to teach Air Force technicians maintenance procedures for a new weapons system. The students complained that the $1M training program left them confused and ill prepared to perform job tasks. If a major vendor with skilled professionals can make such obviously poor design decisions in using advanced training technologies, then the extent of the problem within the Air Force training community could be severe.

This problem in making effective use of new technologies is compounded by the tendency to use more distributed training. Delivering instruction at the job site has the obvious advantage of keeping personnel close to their daily duty functions. The disadvantage is that the job sites lack adequate numbers of skilled and knowledgeable instructional design and development personnel.

The choices are two: 1) provide the needed expertise in the field using tools and expert systems to aid the instructional design and development process, or 2) develop effective training to be delivered independently via computer at the job site. Whichever path is taken, the signpost is obvious: effective use of new training technologies is the key to the solution.

A recent survey at Lowry Technical Training Center revealed that a general lack of knowledge concerning educational principles and their implications for instructional design existed at all levels of the decision-making hierarchy (Muraida & Spector, 1989). Personnel readily admitted that they were not comfortable in performing many of the design and development functions necessary to develop quality instruction.

This problem is compounded by the need to use these new technologies to train increasingly complex weapons systems. The complexity of new systems has two implications: 1) the resulting training requirements will be highly complicated, and 2) the requirement to train effectively increases in importance -- complex new systems are typically placed in situations calling for the highest system reliability. Skill acquisition and skills maintenance become extremely important. The promise is to handle the complex training requirements using new technologies. The problem is to deliver instruction that is highly effective.

The amount of knowledge necessary to design instruction in these situations will continue to increase rapidly. A current instructional design system based on expert system technology might take about 10,000 rules. This is already too many to expect any individual to master. We don't have years to train instructional designers. The problem is that we need instructional design expertise with advanced technologies now, and we need to maintain that knowledge as new techniques and principles are established.

Solution

The problems just discussed indicate a growing need for instructional design expertise, especially for training of complex tasks to be delivered using advanced interactive technologies. However, these technologies are new and cognitive science is embryonic. As a consequence, there is no sizeable cadre of instructional design experts.

The shortage of human experts suggests the possibility of capturing the best human instructional design knowledge in an expert system. Several instructional design expert systems are under construction, the most well-known of which is probably Merrill's ID EXPERT (Merrill & Li, 1989). These systems will most likely prove useful given the complexity of today's training environment, although Richards (1989) argues that instructional designers should proceed with caution in using expert systems. The development of automated design and development tools will also provide a key component in addressing the problem of effectively using these new technologies.

Even if robust instructional design expert systems are implemented successfully, there will be a need for additional research to develop an integrated instructional theory. Encapsulating current knowledge about instructional design in an expert system should not preclude the expansion of our knowledge about instructional design.

A well developed instructional theory should be informed by established learning theories and should consist of specific instructional strategies, principles, and prescriptions which can be empirically refined and refuted. What is needed is a systematic way to test instructional tenets, some of which will pertain to the optimal use of advanced interactive technologies to teach complex activities involving a variety of knowledge types and skills. The test vehicle should provide a means to automate as much of the development and delivery of the instructional materials as possible for the dual purpose of providing rapid development of designs and maintaining test integrity. The Air Force Human Resources Laboratory is designing AIDA to provide the necessary experimental platform.

AIDA is based on the following assumptions: 1) behaviorism is not adequate in providing an explanation of complex learning activities, 2) cognitive science suggests that instruction should be sensitive to the limitations of short term memory, the types, representations, and organization of knowledge, and the situated and integrated nature of human activities, and 3) the principles and prescriptions most in need of empirical test and refinement are those which pertain to the selection, sequencing, and presentation of materials for specific lesson components in a highly automated and individualized learning environment. In short, AIDA is a means for keeping instructiona design apace with advancing educational technology.

Ullmer (1989) has identified three elements of an educational technology: 1) instrumentation, 2) teaching regimen, and 3) image of the learner. With regard to the learner, AIDA will provide a means to determine to what extent and in which circumstances learner control is desirable. With regard to instrumentation, AIDA will provide a means to determine when and how new technologies, such as digitized audio and video, can best be utilized.

The most difficult element to test will most likely be the teaching regimen. Weissman (1988) and others have argued that one of educational technology's main tasks is to provide students with a learning environment that is rich in both tools and data so that learning regimens can be tailored to specific needs.

Conclusion

The current AIDA research plan is guided by these three steps: 1) establish an initial theoretical foundation based on cognitive and behavioral learning theories, 2) formulate strategies, principles, and prescriptions consistent with that theoretical framework, and 3) test, evaluate, and refine these tenets in a variety of instructional settings. The first step will be complete in May, 1990. The second step is expected to be complete in 1991, and testing is expected to begin in 1992.

The challenge that confronts us with these new technologies is to understand better how people learn and to foster those learning processes using new technologies. Without a systematic methodology for testing instructional strategies, principles, and prescriptions, progress is likely to be haphazard and wasteful, especially for complex tasks.

References

Merrill, M. D., & Li, Zhongmin (1989). An instructional design expert system. Journal of Computer-Based Instruction, 16 (3), 95-101.

Muraida, D. M., & Spector, J. M (1989). Needs assessment report: Visit to Lowry AFB Technical Training Center (3400 TCHTW). Brooks AFB, TX: Air Force Human Resources Laboratory.

Montague, W. E. (1989). What works: Summary of findings with implications for Navy instruction and learning. San Diego, CA: Office of Chief of Naval Education and Training.

Montague, W. E., Wulfeck, W. H., & Ellis, J. A. (1983). Quality CBI depends on quality instructional design and quality implementation. Journal of Computer-Based Instruction, 10 (3,4), 90-93.

Richards, B. F. (1989). Should instructional designers design expert systems? Educational Technology Research and Development, 37(3), 63-70.

Ullmer, E. J. (1989). High-tech instructional development: It's the thought that counts. Educational Technology Research and Development, 37(3), 95-101.

Weissman, R. (1988). From the personal computer to the scholar's workstation. Academic Computing, 3(3), 10-41.

Factors Affecting Career Decisions of Minorities in Naval Aviation

Annette G. Baisden
Naval Aerospace Medical Institute
LCDR F. Douglas Holcombe
Naval Safety Center

This poster session reports the development and results of a survey questionnaire designed to identify the reasons minority aviators decide to join, remain in, and leave naval aviation. About one-third of all first-tour minority aviation officers returned the survey. Results suggest that some reasons are important for attracting and retaining individuals, but that different reasons may account for decisions to leave the Navy.

Gender Differences in USAF Pilot Training Performance

Thomas R. Carretta, Ph.D.
Air Force Human Resources Laboratory

The Unites States Air Force (USAF) began to admit women to Undergraduate Pilot Training (UPT) in September 1976. By September 1988, 422 of 637 (66.3%) female pilot candidates had completed UPT successfully compared to 77.7% of the males during the same period (USAF, 1988). The purpose of this study is to determine whether gender differences in UPT performance can be explained by scores on pre-selection instruments (e.g., aptitude test scores, academic grades) currently being used to select pilot candidates. The subjects in this study include 14,240 male and 410 female UPT students who completed UPT since September 1976. Gender effects are being examined separately for each source of commission (Air Force Academy, Reserve Officer Training Corps and Officer Training School) as each source has its own pilot candidate selection procedures. The predictor set includes Air Force Officer Qualifying Test scores and gender by test score interaction terms.

A series of regression analyses are being performed to determine whether there is evidence of differential validity (differences in regression slopes, intercepts or dispersion) when predicting UPT performance (e.g., graduation vs elimination; class standing) for male and female candidates. Implications for USAF pilot candidate selection will be discussed.

References

United States Air Force (December 1988). A Comprehensive Analysis of Gender Neutral Undergraduate Flying Training. HQ USAF/DPXXF, Washington D.C.

1989 Survey of Air Force Life

Louis M. Datko
Air Force Military Personnel Center

Abstract

Every three years since 1977, the USAF Survey Branch has used an omnibus "quality of life" survey to query members Air Force wide. The Survey of Air Force Life (SAFL) is the only Air Force research that longitudinally examines trends of members' perceptions as well as present day issues. Various functional agencies participate in the survey by nominating issues of current concern. Included in the 1989 SAFL was functional representation on overseas policies and programs, housing, Family Support Centers, health resources, public affairs and human relations. Using a stratified random sampling technique, the SAFL, administered to 12,737 Air Force members, provides senior Air Force leadership information by grade-groups, gender, marital status, Total Active Federal Military Service (TAFMS), aeronautical rating, major command and career intent. Trend analysis of the data provides insight into retention patterns, career intent, satisfaction with Air Force life, economic well-being, patriotism and work relations. Further, a detailed approach will explore possible multivariate relationships between satisfaction with the Air Force and career intent for the officer corps. With this data added to that obtained in previous surveys, the Air Force will have documented over a decade of trends on Air Force members' perspectives.

MMPI-2 UPDATE

Dennis Gaynor
5605 Green Circle Drive
Minnetonka, MN 55343
800-627-7271 ext. 5139

The Minnesota Multiphasic Personality Inventory (MMPI) has a long history of application in military surroundings. At this session, there will be a discussion of the restandardized version of the MMPI-2, and handouts of the new scoring reports. The original form of the MMPI was used for over 50 years with no revision in item wording or content.

With the MMPI-2, the traditional features of the original test format remain unchanged:
-the basic scale set
-separate profile forms for male and female subjects
-handscoring keys
-norms with or without K corrections

The revised MMPI-2 features are:
-revised test booklet
-restandardized national norms
-uniform T scores
-new content scales

Also, there will be an Obscan 5 scanner on display. This is the latest in our line of scanners.

Predicting Expert Individual Decision Making Effectiveness

William H. Hendrix
Clemson University

Recently there has been a great deal of interest in developing expert system models. Usually an expert system is based on a single expert, however, policy capturing research suggest that experts frequently differ in their models given the same information or cues. This research investigated within the Brunswick Lens model whether expert effectiveness could be predicted from a rather simple policy capturing project. Results indicated that performance on a policy capturing project was predictive of performance on two expert system type projects. The results suggest that more effective expert system models can be developed by selecting experts based on their performance on policy capturing projects.

Data for Improvement of United States Academies

Clark L. Hosmer
39 Longwood Drive
Shalimar, FL 32579
(904) 651-1739

This is to report a project, "The Clark and Lyn Hosmer Educational Memorial Trust," for an annual competition among the five United States Academies of the Air Force, Army, Coast Guard, Merchant Marine, and Navy.

Our purpose is to provide data on the elements of schooling that contribute to graduates' effectiveness on job and in the community. The purposes of the academies and their Associations of Graduates (AOG) include production of effective graduates, large class memberships in the AOG, and collection of contributions for the Academies' betterment. Our project will help to fulfill the academies' and our purposes, at minimum cost.

Part of the low cost of the plan derives from concentration on a small portion of the elements of schooling. The data to be collected are only the elements, if enlarged or improved, that would pay off most in graduates' effectiveness. Those are the elements at the top and the bottom of a theoretical list of <u>all</u> the elements ranked in order of contribution to graduates' effectiveness. The Services use the same principle in evaluating the graduates during their careers. Those at the top are 'enlarged' by promotion; those at the bottom, unless improved, are not and may be separated.

The Trust calls for the AOG's in their alumni publications to request each graduate in the classes 5, 10, and 20 years out to name only the two elements of schooling that contributed most and least to his or her effectiveness on the job and in the community. After several years' experience, the AOG's request could list the major categories that had been most frequently cited, plus an 'other' alternative. At the present time, we can only guess what the data might be. The actual data will be the payoff as one of the bases for strategic review of the conduct of the performance of the mission of the Academy.

The Academy with the highest number of usable returns before President's Day from the members of the designated classes, divided by the three classes' total graduates minus known deceased, will win the annual competition. That Academy will have its name inscribed on and retain custody of the Trust's 100 year trophy, until a different Academy wins.

The class of the winning Academy that contributes the highest percentage to the total percentage, will win for its Class Fund at least $1,000.

The Trustees will insure that the cash awards will continue to have at least the same value by retaining from income and adding to the Trust principal each year the rate of inflation times the principal.

Your criticisms are <u>welcome</u>. Thank you.

The Theory of Survival Psychology

Captain Robert I. Lawconnell, PhD
Frank J. Seiler Research Laboratory
USAF Academy CO 80840–6528

Abstract

A utility function which may be useful in predicting human behavior and decisions is derived. It is based on the premise that the primary purpose of an individual/family/group is survival. Survival is assumed to depend on obtaining at least the minimum required amount of needed resources. It is also presumed that obtaining more than the minimum becomes less and less important in accordance with the excess. The derived utility function is then used to make theoretical predictions for a variety of repetitive and non–repetitive choice scenarios. The theoretical results are compared to the corresponding experimental data and are seen to be in good agreement. It is also interesting to note that many aspects of risk aversion and risk taking follow naturally from the utility function.

Effects of Spatial Filtering on Visual Search and Target Detection

Major George R. Mastroianni, Ph.D. and
Harry Zwick, Ph.D., Letterman Army Institute of Research

Cadet Bryan Campbell, United States Military Academy

Abstract

It has been shown that exposure to low-level laser radiation, such as that from rangefinders or target designators, can cause temporary alterations in spatial sensitivity. The aim of this research program is to quantify the effect of spatial filtering on the detection of military vehicles in photographic images. Predictions of the magnitude of expected gunner performance decrements resulting from these studies will contribute to efforts to model the effects of directed energy warfare. Basic research efforts conducted as part of this program are intended to characterize the role of spatial vision in visual search and target detection, and will provide essential theoretical support for such problems as the design of effective camouflage and the potential integration of near real-time image processing into visual sensors.

Photographs of armored vehicles in fighting positions in West Germany are used as stimuli. These images are subjected to two-dimensional fast-Fourier transformation. The images are filtered to pass only high frequencies, only low frequencies, or to mimic the degradation in spatial sensitivity observed in monkeys after exposure to lasers. These images can be produced either in color or in black and white; overall contrast and scene luminance can be closely specified. The stimuli are presented using a computer projection system that controls the timing and duration of the trials.

Current studies use a simple psychophysical procedure to determine the effect of filtering on detection latency and accuracy. Performance is compared across all four stimulus conditions - the three filtered images and the original, natural scene. We are in the process of integrating a SRI Generation V dual-Purkinje-image eye tracker into the experiment to study oculomotor aspects of visual search. Behavioral contingencies will eventually be overlaid on the basic detection paradigm to assess more realistic cognitive interactions with visual search of degraded scenes.

Effect of a Simulated Sustained Flight Operation on Cognitive Performance, Grip Strength, and Mood

David F. Neri, LT, MSC, USN
Scott A. Shappell, LT, MSC, USNR
Charles A. DeJohn, D.O.
Naval Aerospace Medical Research Laboratory
Dennis L. Reeves, LCDR, MSC, USN
George Washington University

Abstract

Sustained flight operations (SUSOPS) tend to produce both mental and physical fatigue, as well as changes in mood. This study assessed changes in auditory and visual performance, grip strength, and subjective mood during a simulated SUSOP. The scenario consisted of a 9-hour planning session followed by 4 hours rest and a 14-hour nighttime mission (1900-0900). After 6 hours sleep, the 9-4-14 work-rest-work pattern was repeated, with a daytime mission occurring from 0400 to 1800. Several times during the planning and mission segments, 12 subjects performed a battery of computer-administered tasks, including the Dichotic Listening Test (DLT) and a sonar detection task. Grip strength was measured repeatedly with a dynamometer and subjective mood with the Addiction Research Center Inventory (ARCI). According to an analysis of variance, subjects showed a significant increase in the number of errors on the DLT, a significant drop in correct detections on the sonar task, and a significant reduction in grip strength during both 14-hour missions. The ARCI revealed that subjects felt an increase in fatigue during the nighttime mission, with little change during the daytime mission. They also reported a decrease in intellectual efficiency during the nighttime mission and a slight increase during the daytime one. We conclude that as few as two consecutive 14-hour missions can result in a decline in performance as well as self-reported fatigue. No performance decline was obtained on a visual search or tracking task. Subsequent analyses will compare these data to a scenario phase-shifted by 12 hours, to further examine time-of-day effects.

The Effects of Uncertainty of Nap and Work Duration on Cognitive Performance and Mood State During Sustained Operations

Robert J. Pleban, Ph.D.
U.S. Army Research Institute
for the Behavioral and Social Sciences
Fort Benning, Georgia

Barbara J. Ross, M.S.
Department of Behavioral Biology
Walter Reed Army Institute of Research
Wahington, D.C.

Abstract

Research was performed to examine the cognitive and subjective effects of uncertainty of nap and work duration during sustained operations. Eighteen male subjects were assigned to one of three conditions: uncertain nap, certain nap, and no nap. The uncertain group differed from the other groups in that subjects were not given any specific information on work-sleep cycling. With the exception of a three hour nap period, the work-sleep schedules were identical for each group. Performance and mood were assessed over three sessions (one 12 hour baseline session and two 17 hour work sessions) using a variety of computer based tasks. The results showed that uncertainty of work-sleep duration had both positive and negative effects on cognitive performance depending on the specific task that was employed. Reported changes in mood state across time were generally small with no consistent group pattern. Overall, the findings underscore the value of naps as a means of extending performance effectiveness during sustained operations. More specifically, these results indicated that for commanders and those individuals whose mental workload is extensive, some type of sleep discipline must be practiced. While the short term impact of napping may be negative (i. e., sleep inertia effects), the long term performance sustaining qualities of even low quality sleep can be significant when compared to no sleep at all.

Prediction of Training Success in the

Special Forces Qualification Course

Robert J. Pleban, Ph.D.
Thomas J. Thompson, Ph.D.
U.S. Army Research Institute
for the Behavioral and Social Sciences
Fort Benning, Georgia

Abstract

Recent research is reported which examined the predictive utility of selected measures of intelligence (Wonderlic Personnel Test - WPT), personality (Jackson Personality Inventory - JPI), and biographical information in predicting the successful completion of the Special Forces Qualification Course (SFQC). Intelligence, personality, and biographical measures were obtained from 293 soldiers attending Phase I of the SFQC at Fort Bragg, NC. The relationships between two of the selected predictor measures, WPT and JPI, and the Phase I performance criteria were assessed through a correlation-multiple regression strategy. Predictive utility was assessed using a discriminant analysis procedure. Intelligence (WPT) correlated significantly with overall Phase I status. Significant correlations were also obtained between Phase I status and four JPI scales: Energy Level, Anxiety, Risk Taking, and Infrequency. The results further suggested that SF candidates who have had prior specialized training emphasizing land navigation, map reading, and patrolling (e.g., Ranger, Reconnaissance, Jungle Warfare) are likely to be much better prepared to complete Phase I successfully than those candidates who have not had such training. The single best predictor of Phase I status was intelligence. The best predictive model of Phase I status consisted of three variables: the WPT, Energy Level, and Risk Taking. When employed with the present sample, this model was able to increase the total percentage of correct classifications (successful-unsuccessful) by 11.5%. It was concluded that the use of a valid screening program could offer a significant reduction in training costs by insuring that the selected candidates have the basic capabilities and temperament for mastering the content of the course, thus minimizing attrition and the number of candidates recycled.

Field Observation Methods Training For Military Behavioural Scientists

Gerald D. Resch
P.R.A. Personal Applied Research Associates
Victoria British Columbia

Summary - Poster Session

Within the Canadian Forces over one hundred officers from the Personnel Selection Branch are employed as personnel specialists and behavioural science advisors to bases and units. Increasingly, their responsibilities include the conduct of field research in the Sea, Land and Air environments, as well as the provision of behavioural science advice to field unit commanders. In an effort to formalize training in Field Observation Methods, the Director of Personnel Psychology and Sociology (DPPS) at National Defence Headquarters, Ottawa contracted the development and conduct of a Field Observation Methods course. The course was conducted during a Brigade exercise at Hohenfells, Germany, from 21 August to 10 September, 1989. Each of the 14 participants attended a one week theory and orientation session in Lahr, Germany, and then joined separate combat and combat service support units for two weeks in the field.

The course material brought together theoretical and practical findings from the fields of Social Psychology and Sociology. All material was adapted for application to the military field setting based upon the experience of the Canadian Service Women In Non Traditional Environments And Roles (SWINTER) Land Trials and relevant allied military studies.

All methods of field observation, from complete participation to complete observation were reviewed. Emphasis, however, was given to the roles of participant-as-observer (PO) and observer-as-participant (OP). An active rather than passive role was advocated and skill in adopting this role without interfering with the natural setting was imparted.

Human relations skills as well as the survival and facilitative skills required for effective observation in combat arms units were discussed. Topics included:

- strategies for reconnaissance, contact, negotiation and entry;
- issues of reliability and validity of field observations;
- methods of recording, calibration of observers, debriefing and interpretation;
- field research development and report writing; and,
- ethical considerations.

The course was successfully conducted at the level of post-graduate work in the behavioural sciences. Course members were deployed as a research team to investigate practical methods of observing unit cohesion and to produce a behavioural science needs analysis for brigade units. The course members were required to produce a final report of the exercise prior to leaving Europe. This approach was adopted for the training of all future field observers/advisors for the Canadian Forces.

Teacher's Perceptions of Severity Ratings and the Need to Report Maltreatment in Military Operated Schools

JanaLee L. Sponberg, DAEd

Child maltreatment, comprised of child abuse and neglect, is a serious problem. Maltreatment situations cover a very broad range of actions and failures to act, differing greatly in their nature and severity. The effects of maltreatment may range from major to minor and may be physical injury, psychological trauma, sexual maltreatment, deprivation of necessities, or a combination. Child maltreatment is a complex issue involving families in which children are suffering maltreatment and involving people who are attempting to protect these children. Professionals such as social workers and educators need to be communicating, cooperating, or collaborating to protect children. This poster session will present research findings examining areas of similarity and differences in responses to maltreatment severity by teachers and social workers.

DoD Child and Spouse Abuse Data

JanaLee L. Sponberg, DAEd

The office of family policy and support, a directorate under the deputy assistant secretary for defense for family support education and safety, is responsible for coordination and oversight of the family advocacy program and related issues. DoD instruction 6400.2, "Child and Spouse Abuse Reports" July 10, 1987, provides guidance for DoD reporting requirements.

Information collected by OFP&S between FY86 and FY89 will be available to be presented to an audience through the graphic samples attached and handouts. The materials will include the following:

Annual Report for FY89

Trend Analysis - FY86-FY89

Statistical Comparison

Service FAP Information Handouts

Two studies are currently under contract with the American Association for Protecting Children to examine and refine the Departments child and spouse abuse data collection effort. Material could be made available for formal presentation of these projects.

Personality-Gender Interactions in the Workplace

Laurie C. Walters, Ph.D.
Air Force Human Resources Laboratory

As more females assume leadership positions in traditionally "masculine" career fields, it is important to consider how the female leader is accepted by both subordinates and superordinates. Individual differences may significantly affect how leaders are perceived. The present study examined how subordinates and superordinates, who vary on authoritarianism, evaluate male and female leaders.

Subordination to a female leader presents a "dilemma" for the male authoritarian, as various characteristics of the authoritarian personality imply rejection of the female, while others imply acceptance. Authoritarians, who are somewhat rigid and stereotypical in their thinking, may reject female leaders who are perceived as violating conventional norms. On the other hand, authoritarians also tend to be submissive and reverent towards authority figures. As a result, they may actually evaluate the female leader favorably. However, when no longer in a subordinate position, authoritarians may be more negative in their ratings of female leaders.

Subjects were 233 male college students who, after completing a measure of authoritarianism, assumed the role of subordinate or superordinate as they watched a video of a male or female leader conducting a training group. Subjects then evaluated the leader on such attributes as competency, likeability, promotability, etc.

Results indicated that authoritarians, regardless of their position, were significantly more favorable towards male leaders than female leaders. Low authoritarians, however, tended to be more favorable towards female leaders than male leaders when a preference was indicated. The interaction of gender and personality suggests possible impacts on mixed-gender crew coordination.

ASVAB and Demographic Profiles of Enlisted Navy Males in a Biodynamics Research Program: A Comparison of Program Completers and Noncompleters

LT Schuyler C. Webb, MSC, USNR
Naval Biodynamics Laboratory
New Orleans, LA

Abstract

The Naval Biodynamics Laboratory uses junior enlisted Navy men as Human Research Volunteers (HRVs) in its Aircrew Injury Prevention and Human Factors research program. As the need for HRVs increases and qualifying examination costs rise, optimal HRV selection is critical. As a result, researchers have paid increasing attention to differences between HRVs who successfully complete their research assignment and HRVs who do not. The Armed Services Vocational Aptitude Battery (ASVAB) is required of all incoming enlisted personnel. It consists of subtests designed to measure aptitudes predictive of success in military occupations. The purpose of this study is to evaluate ASVAB subtest scores and demographic profiles as predictors of HRV success in our research program.

Method

ASVAB subtest scores and demographic profiles of two HRV groups were compared. The successful completion of their research assignment was the major criterion for evaluation.

Results

One group (N=16) completed the research assignment; the second (N=12) did not. Completers scored significantly higher ($p<.01$) on the Word Knowledge subtest than noncompleters. Correlations between Word Knowledge and Mechanical Comprehension scores and program completion were positive ($r=.44$ and $.32$, respectively). Both groups were equivalent across demographic measures. However, completers had a higher educational level and enrolled part time in college during their tour.

Conclusion

The Word Knowledge score is the best predictor of HRV research program completion. Concomitant with other evaluative measures (i.e., personality, motivation, etc), it can be a useful supplementary predictor of HRV success in biodynamics research participation. Future studies will investigate other cost efficient measures to predict HRV assignment outcome.

The Enduring Effects of Vietnam on Political Attitudes

Minda Weston and Michael D. Matthews
Drury College

Charles N. Weaver
St. Mary's University

Abstract

This poster summarizes data examining the political attitudes of persons who were draft age during the Vietnam conflict (cohort I) and older persons who were not draft age during the conflict (cohort II). The data for this study were derived from the General Social Surveys, conducted annually (except for 1979 and 1981) since 1972 by the National Opinion Research Center. Cohort I was defined as those persons who were ages 18-25 in 1973, and cohort II was defined as those persons ages 36-43 in 1973. Responses of males and females were compared for surveys administered from 1973 through 1988. Responses to four questions were examined. These questions focused on military spending, foreign aid spending, the likelihood that the U.S. will be involved in another major war within the next 10 years, and whether the U.S. should take an active role in world affairs. The results indicated that, over the years studied, cohort I believed we are spending too much on defense. Similar results were obtained with respect to spending on foreign aid, except the differences between the two groups diminished through the 1980s. Cohort I also consistently expressed a stronger belief that the U.S. will be involved in another war soon. With respect to its role in international affairs, the results are more complex, with there being an interaction between cohort and year of survey. For all questions, responses of males and females were similar. Implications of the results will be discussed.

★ U.S. GOVERNMENT PRINTING OFFICE: 1990—774-072