



DTIC FILE COPY

AD-A193 194

# Independent Research and Independent Exploratory Development Programs

## FY 87 Annual Report

DTIC  
ELECTE  
MAY 06 1988  
S D  
CD



GENE SCOUT

Approved for public release; distribution is unlimited.

88 5 05 009

## Foreword

Independent Research (IR) and Independent Exploratory Development (IED) funds are provided to the Technical Directors of Navy Laboratories as discretionary funds to support innovative, promising research and development outside the procedures required under normal funding authorization. The funds are to encourage creative efforts important to mission accomplishment. They enable promising researchers to spend a portion of their time on examining the feasibility of self-generated new ideas and scientific advances. They can provide important and rapid test of promising new technology, and can help fill gaps in the research and development program. This may involve preliminary work on speculative solutions too risky to be funded from existing programs.

The funds also serve as means to maintain and increase the necessary technology base skill levels and build in-house expertise in areas likely to become important in the future. These programs contribute to the scientific base for future improvements in the manpower, personnel, and training system technology.

Research at the Navy Personnel Research and Development Center addresses the Navy's needs for enhancing system and personnel performance through the integration of people and technology. Resources provided for the IR/IED program have been used to develop a variety of research methods, models, and techniques within the areas of training, manpower utilization, organizational productivity, and human factors engineering of naval weapon systems and platforms.

The IR program has been active at this Center since 1973 and is funded under Program Element (PE) 0601152N. The IED program was initiated in 1976 and is funded by PE 0602936N.

The IR/IED programs for this reporting period are shown in Tables 1 and 2. The projects are described in detail in subsequent sections of the report followed by appendices containing additional information relevant to the IR/IED program.

**B. E. BACON**  
Captain, U.S. Navy  
Commanding Officer

**J. S. McMICHAEL**  
Technical Director

# Independent Research and Independent Exploratory Development Programs

## FY87 Annual Report

*William E. Montague*  
(Editor)

Approved by  
*J. S. McMichael*  
Technical Director

Accession For	
NTIS CRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By _____	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
<b>A-1</b>	

Approved for public release; distribution is unlimited.



**Navy Personnel Research and Development Center**  
San Diego, CA 92152-6800

IR/IED FY87 Annual Report

**Table 1**  
**Independent Research**  
**Work Units for FY87 and FY88**  
**(PE 0601152N)**

Work Unit	Title	Principal Investigator	Internal Code	Telephone (619) 553- or A/V 933	FY Funding (\$K)	
					87	88
R000-N0-000-						
01	Brain Mechanisms on human color vision: Implications for display systems	Trejo/Lewis	41	7981	55	55
02	How to elicit knowledge from experts	Bamber	41	9219	0	40
03	Stabilization of performance on a complex cognitive task	Federico	51	7688	65	60
04	Experienced-based career development	Morrison	62	9256	0	41
05	Event-related potential correlates of memory performance	Williams	41	7925	0	35
RR000-01-042-04-						
024	Models for calibrating multiple-choice items	Sympson	63	7610	40 <sup>a</sup>	0
025	Policy modeling techniques for large-scale multiple objective problems	Liang	61	7959	80	0
RR000-01-042-06-						
026	Construct validity of instruments for appraising performance	Kidder	62	7632	40	0
					<u>280</u>	<u>231</u>

<sup>a</sup> Transitioned to IED

IR/IED FY87 Annual Report

Table 2

Independent Exploratory Development  
Work Units for FY87 and FY88

Work Unit	Title	Principal Investigator	Internal Code	Telephone (619) 553- or A/V 933	FY Funding (\$K)	
					87	88
RV36-I27-						
01	Optimal control theory for a system of quasi-linear difference equations	Krass	612	7962	0	50
02	Reading comprehension strategies evaluation	Baker	522	7305	0	30
03	Models for calibrating multiple choice items	Sympson	63	7610	0 <sup>a</sup>	40
04	Organizational simulation laboratory: IR/IED	Nebeker	41	7749	60	50
RF66-512-						
018	Trend analysis for real-time stochastic problems	Malkoff			35 <sup>b</sup>	0
019	Changes in cognitive structures with training	Montague	51	7849	50 <sup>b</sup>	0
020	Statistical process control	Landau	42	7937	30 <sup>b</sup>	0
WR-48468	Artificial intelligence tryout center	Montague	51	7849	12	0
					<u>187</u>	<u>170</u>

<sup>a</sup> Transitioned from independent research

<sup>b</sup> Research completed

# CONTENTS

	Page
<b>Independent Research (IR) Program</b>	
<b>STABILIZATION OF PERFORMANCE ON A COMPUTER-BASED SIMULATION OF A COMPLEX COGNITIVE TASK</b> Pat-Anthony Federico	1
<b>PERFORMANCE APPRAISALS: CAN WE DISCOVER "WHAT" WE ARE MEASURING</b> Pamela Kidder	6
<b>POLICY MODELING TECHNIQUES FOR LARGE SCALE MULTIPLE OBJECTIVE PROBLEMS</b> Timothy Liang	10
<b>MODELS FOR CALIBRATING MULTIPLE-CHOICE ITEMS</b> James Bradford Simpson	13
<b>BRAIN MECHANISMS FOR HUMAN COLOR VISION-- IMPLICATIONS FOR DISPLAY SYSTEMS</b> Leonard J. Trejo and Gregory W. Lewis	19
<b>Independent Exploratory Development (IED) Program</b>	
<b>STATISTICAL PROCESS CONTROL AS AN ENHANCEMENT TO JOB AND ORGANIZATIONAL DESIGN</b> Samuel B. Landau	30
<b>THE EFFECTS OF GOALS, STANDARDS, AND REWARDS ON WORK PRODUCTIVITY AND QUALITY</b> D. M. Nebeker, B. C. Tatum, and B. L. Cooper	35
<b>EFFECTS OF PRACTICING QUALITATIVE PROBLEMS ON TEST PERFORMANCE IN A BASIC ELECTRICITY COURSE</b> William E. Montague	47
<b>APPENDIX A--PROJECT TRANSITIONS</b>	A-0
<b>APPENDIX B--PRESENTATIONS AND PUBLICATIONS</b>	B-0
<b>APPENDIX C--HONORS AND AWARDS</b>	C-0
<b>DISTRIBUTION LIST</b>	



# STABILIZATION OF PERFORMANCE ON A COMPUTER-BASED SIMULATION OF A COMPLEX COGNITIVE TASK

Pat-Anthony Federico

---

*The purpose of this research is to study the processes intrinsic to the stabilization of performance on a complex cognitive task, (conducting an outer air battle). Subjects will interact with an animated, computer-based, graphic simulation. They will allocate, deploy, and manage tactical assets in a very large number of scenarios to defend carrier-based task forces against hostile, missile-launching bombers. Concurrent and retrospective verbal protocols will be obtained from the subjects regarding their battle management. Performance during each scenario will be automatically assessed by the computer system against 16 multivariate measures. Cognitive and statistical analyses will be conducted to study the processes of acquiring skill and reaching stabilization of performance on this complicated mental task. Contributions to methodology and theory culminating from this research will result in improved operationally oriented performance assessment.*

---

## Background

Individuals vary in their rates and manners of skill acquisition especially in the beginning of practice, and they reach terminal performance plateaus differentially. Early performance requires high conscious control (i.e., it is slow, sequential, effortful, limited, and directed), whereas late performance tends to be automatic (i.e., it is fast, parallel, effortless, and less limited by attentional focus). Practice during the early stages results in dramatic changes in

behavior (e.g., decreasing performance variability, minimizing response time). With practice, rate of improvement diminishes and becomes more uniform across individuals (i.e., performance stabilizes). For some tasks, performance does not seem to get any better or worse, and curves that reflect the rate of skill acquisition of individuals appear to be parallel (Ackerman & Schneider, 1984; Jones, 1984; Schneider, 1984). Individual variability among learners affects modes and speed of skill acquisition:

Distinct experiences, cognitive models, aptitudes, and motivation can influence early and late performance differentially.

Much of the earlier research on which the above statements are based was done with psychomotor tasks. A lot less is known about complex tasks, which are primarily cognitive in nature.

### Problem

Because many factors affect the nature and time course of acquisition, beginning performance on complicated tasks is usually not a good estimate of terminal performance. Estimates of performance are likely to measure different things on different trials for different people. Trying to separate accurately better and poorer performing people, or to determine consistently whether a trainee has mastered a needed skill become difficult. This potential lack of reliability impacts upon the predictive power of computer-based simulations for assessing operationally oriented skills. Therefore, it affects the validity of computer simulations for job-sample-performance testing in functional contexts.

### Technological Objective

The technological objective of this proposed research is to conduct cognitive and statistical analyses as well as theoretical modeling to study the process of skill acquisition resulting in the stabilization of performance on a computer-based simulation of a complex cognitive task.

## General Approach

### Target Task

The target task of this proposed research consists of tactically allocating, deploying, and managing fighter and supporting aircraft to defend an aircraft carrier and its escorting ships against threatening Soviet naval air bombers. This task demands considerable practice before it can be executed with a sufficiently high level of skill and becomes automatic. For the purposes of this research, this task is considered as a test of individual differences in complex mental performance. In the execution of this task, the transition from controlled to automatic performance is important. This implies that what is crucial is not early but late performance (i.e., how well individuals do after extended practice). The administration of numerous trials on this task, together with cognitive and statistical analyses, will make it possible to note when and how stabilization of performance is achieved (i.e., when the research subjects no longer show any tendency to improve or worsen with practice).

### Computer-Based Simulation

Software tools were developed for constructing computer-based animated graphic simulations of the actual radar coverage of F-14 and F/A-18 fighters and E2-C early warning aircraft as well as fuel flow of these planes together with KA-6 tankers. These include probability of kill for Phoenix, Sparrow, and Sidewinder missiles that the different fighters carry as well. The capabilities

to generate an infinite number of raids from Soviet naval air bombers with antiship missiles (ASMs) in different warfare theaters and various carrier loadouts in terms of numbers of each type of fighter and missile on board enable the creation of an infinite set or universe of tactical scenarios. These will be used to assess how well individuals manage outer air battles to defend carrier-based naval task forces.

### Subjects

The research subjects, approximately six F-14 pilots and radar intercept officers at NAS Miramar and/or instructors and students from the Tactical Action Officer, Tactical Warfare Overview, and/or Staff Tactical Watch Officer Courses from the Fleet Combat Training Center Pacific, will be required to allocate, deploy, and manage fighter and supporting aircraft in order to knock down various numbers and mixes of hostile bombers before they reach their respective ASM launch points. Each computer-based scenario will be run in compressed or accelerated time; each threat scenario will be considered as a performance test item.

### Performance Criteria

A subject's tactical performance during simulated air battles will be assessed according to 16 multivariate criteria. Some of these are as follows: the percentage of incoming threat aircraft detected by F-14, F/A-18 and E2-C radar systems, the percentage of bombers that fighters placed in missile launch acceptability regions (LARS),

the percentage of hostile aircraft shot down or probably killed, the average range from the defended task force at which threat aircraft were knocked down, the percentage of hostile platforms knocked down before ASMs were launched, etc.

### Procedure

Subjects will be run on the computer-based scenarios of these symbolically displayed air battles between Soviet bombers and U.S. carrier-based aircraft. How well each allocates, deploys, and manages fighters and other supporting aircraft during the simulated battle will be assessed according to the performance criteria mentioned above. The possible number of incoming raids or specific threat scenarios form a practically infinite universe. Consequently, the set of simulated tactical scenarios will be considered as an operationally oriented, domain-referenced, job-sample, performance test. With each scenario as an assessment trial, subjects will be administered 200 trials divided into 20 blocks.

### Cognitive Analysis

During the first trial of every block, verbal protocols will be obtained from the subjects as they are conducting the simulated air battles. The analyses of these verbalizations, as well as retrospective reports, will disclose the information heeded by the subjects while they perform this complex task. Comparisons of the thinking-aloud protocols and retrospective reports on the first trial of every block will reveal

the variability in cognitive processing within as well as between subjects as they acquire skill (i.e., progress from controlled to more automatic performance of the task).

Analysis of protocols obtained early and late during practice on the task will indicate how subjects' cognitive processes and structures change as their performances tend to stabilize. These will reflect the cognitive correlates of the acquisition of stable task performance. Together with a thorough componential analysis, the information obtained from the protocol analysis will be used to construct a model for performing this complex task. This model will be used to create a theoretical framework as well as serve as the basis for programming an expert system: an "intelligent tactician" that will monitor, diagnose, and assess the conduct of simulated air battles to defend carrier task forces.

### Statistical Analyses

Combining statistical procedures with protocol analyses and conceptual modeling will provide an integrated account of the cognition accompanying the acquisition of complex task performance. Together with cognitive analysis and theory, statistical techniques (e.g., a test for the homogeneity of  $k$  regression lines) can be used to uncover the mental processes and structures underlying the acquisition of stabilization.

### Potential Products/Transition

The potential products of this research are contributions to a knowledge base and much needed

theoretical framework. The methodology and theory culminating from this research can be extended or transitioned to the exploratory development of "intelligent or expert" computer-based simulation systems to measure complex cognitive performance in functional contexts. Then, the predictive power of this type of performance assessment can be determined. Likewise, this follow-on work itself can be transitioned to advanced development of an intelligent computer-based simulation system to support job-sample performance assessment of intricate cognitive tasks. This advanced system would allow the accessing of developed methodologies, theoretical orientations, mental models, as well as generic software tools to implement prescriptive procedures to aid in the production of performance tests for complex cognitive tasks.

### References

- Ackerman, P., & Schneider, W. (1984). *Individual differences in automatic and controlled information processing* (Rep. HARL-ONR-8401). Champaign IL: University of Illinois, Department of Psychology.
- Jones, M. (1984). Video games as psychological tests. *Simulation & Games, 15*, 131-157.
- Schneider, W. (1984). *Training high performance skills: Fallacies and guidelines* (Rep. HARL-ONR-8301). Champaign IL: University of Illinois, Department of Psychology.

## Biography



**PAT-ANTHONY FEDERICO** is a Senior Research Psychologist in the Training Technology Department at NPRDC. He earned his BA *cum laude* from the University of St. Thomas in 1965 with a double major in mathematics and philosophy and a minor in physics. He was awarded his Ph.D. in 1969 from Tulane University in general experimental psychology. He has research interests in individual differences in cognitive processing, learning, and performance; and computer-based instruction and performance assessment. He was

elected and served as Executive Director, President, and Secretary-Treasurer of the Human Factors Society, San Diego Chapter. He is also a member of the Cognitive Science Society, Psychonomic Society, American Educational Research Association, and American Psychological Association. He is a member of the editorial advisory review board for the *Journal of Educational Psychology*, and an ad hoc reviewer for *Human Factors and Memory and Cognition*. He is a peer reviewer and advisor for the Office of the Assistant Secretary for Educational Research and Improvement, United States Department of Education. He has authored or edited 80 scientific contributions including books, chapters, journal articles, professional papers, and technical reports.

## PERFORMANCE APPRAISALS: CAN WE DISCOVER "WHAT" WE ARE MEASURING?

Pamela Kidder

---

*In order to promote, demote, and retain employees, many organizations, including the Navy, have developed performance appraisal systems. Many of these systems are extremely complex. Unfortunately, most performance appraisal systems are ambiguous. No one knows "what" the instrument is actually designed to measure.*

*This research was conducted to discover what performance appraisal systems really measure. The underlying constructs represented by a specific battery of performance appraisal measures were found and identified. A replication of the study is suggested, using a larger sample.*

---

### Problem

Performance appraisals are utilized in most organizations today, including the Navy. Complex instruments have been developed in an attempt to measure and interpret job performance. Unfortunately, many instruments have no solid basis for usage; that is, the performance appraisal methods are seldom validated or tested to determine their effectiveness. This situation is of great concern because many decisions are based on performance appraisal information. Performance appraisals often control the flow of personnel within an organization. Individuals may be promoted, demoted, and/or transferred based on performance appraisals. Appraisals can also provide feedback information for counseling and motivating

participants (Landy & Farr, 1983; Latham & Wexley, 1981). Prior to using performance appraisal instruments for important organizational decision-making, the instruments should be validated. The present study is a preliminary effort in determining the accuracy and usefulness of a performance appraisal system.

### Background

The accuracy of a performance appraisal system must be assessed by evaluating the validity of the system. In general, validity can be assessed via criterion-related validity, content validity, or construct validity. Criterion validity is not a viable alternative in performance appraisal research because the computation of criterion validity requires a

performance appraisal instrument and a closer to perfect performance measure. If this improved measure was known to exist, there would be no need to prove criterion related validity (Kane & Lawler, 1979). Content validity is also inappropriate because it is not really validity; it is a method of test construction or a method of determining the content representativeness of a test (Messick, 1975; Tenopyr, 1977).

An alternative is the use of construct validity, which attempts to determine "what" the instrument measures. This type of validity has been discussed in terms of selection (Campbell, 1976; Guion, 1965; James, 1973) and has been demonstrated in this area. In addition, construct validity has been addressed in the performance appraisal literature (Kane & Lawler, 1979); however, there have been few/no prior attempts to determine the construct validity of performance appraisal systems. Therefore, it is not surprising that there is no clear methodology for assessing the construct validity of appraisal systems. The present study focusses on construct validity and "what" performance appraisal instruments measure.

### Approach

The research was conducted in three primary phases. In the first phase, participants were selected and their job was examined. Electronics Technician (ET) Chiefs (E-7 to E-9) served as participants. The ET supervisor position was chosen because it involves a variety of job duties, including technical,

administrative, and managerial skills; it is a critical position for all branches of the Armed Services. Following the selection of the participants, job analysis data were gathered for the incumbents. The researcher learned as much as possible about the ET Chief job and their performance appraisal system.

In the second phase, instruments were developed. First, a questionnaire was developed. All survey questions were tailored to the ET rating. The survey was designed to gather additional data about the current performance appraisal system and the ideal performance appraisal system, as viewed by ET Chiefs. Second, several sample performance appraisal measures were developed, including Behaviorally Anchored Rating Scales (BARS), a structured performance appraisal interview, and an assessment center exercise.

In the third phase, data were collected. Twenty ET Chiefs and their supervisors participated in the BARS, interview, and assessment center exercise. Thus, performance appraisal data were obtained for all participants on all measures. The performance data were analyzed to determine evidence of construct validity; that is, to determine the underlying constructs that are measured.

### Results and Discussion

The performance appraisal data suggest that underlying constructs can be identified in performance appraisal instruments. Two constructs were found. These constructs were labelled communication and perseverance.

Communication is interpersonal communication and written correspondence. Perseverance is the Chiefs' ability to initiate all action and to follow the action to culmination. The ability to identify the factors is extremely important. The results are promising because they suggest that constructs can be identified in performance appraisal. One must note that this study consisted of a small number of participants. A larger sample should be examined prior to making global statements regarding construct validity.

Only when we discover "what" we are actually measuring can we develop performance appraisal measures that measure what we want to measure. If we can identify constructs, then we can improve the quality of the workforce. In addition, a great savings can be realized through improved promotion, demotion, and retention of qualified personnel.

### References

- Campbell, J. P. (1976). Psychometric theory. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology*. Chicago: Rand-McNally.
- Guion, R. M. (1965). *Personnel testing*. New York: McGraw Hill.
- James, L. R. (1973). Criterion models and construct validity for criteria. *Psychological Bulletin*, 80, 75-83.
- Kane, J. S., & Lawler, E. E. (1979). Performance appraisal effectiveness: Its assessment and determinants. *Research in Organizational Behavior*, 1, 425-478). JAI Press.
- Landy, F. J., & Farr, J. L. (1983). *The measurement of work performance: Methods, theory, and applications*. New York: Academic Press.
- Latham, G. P., & Wexley, K. N. (1981). *Increasing productivity through performance appraisal*. Reading, MA: Addison-Wesley Publishing Co.
- Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. *American Psychologist*, 955-966.
- Tenopyr, M. (1977). Content-construct confusion. *Personnel Psychology*, 30, 47-54.



## Biography



**PAMELA KIDDER** is a Personnel Research Psychologist at the Navy Personnel Research and Development Center. Her current research focusses upon performance appraisals and job performance measurement. Her graduate education was at the University of Maryland in Industrial/Organizational Psychology. She is a member of the Academy of Management, American Psychological Association, American Psychological Association Division 14, American Psychological Association Division 19, and the Western Psychological Association.

# POLICY MODELING TECHNIQUES FOR LARGE-SCALE MULTIPLE OBJECTIVE PROBLEMS

Timothy Liang

---

*The Navy's personnel assignment model matches people to jobs in accordance with multiple policy objectives. It is designed for weekly operation in each rating or occupational specialty. Because the model focuses on the details of operations, it does not have a capability of measuring the impact of multiple policy objectives for a whole year or for a group of ratings. This effort developed a technique to formulate a policy model that links aggregate policy plans to disaggregate operational decisions.*

---

## Background and Problem

The Navy's assignment problem has been recently formulated as a multiple objective transshipment model. The model for some ratings is currently being installed at the Naval Military Personnel Command to test its feasibility in assigning enlisted personnel. Application of the method has been extended to include more ratings. The advantage of using the assignment model to replace the current manual process is not limited to its efficiency in terms of speed and accuracy. More important, it provides a systematic procedure for executing multiple policies. Decision makers may specify the priority of the policies and obtain a set of people-job matches in

accordance with those policies. To meet the current operational need, the assignment model is designed to match people to jobs on a daily or weekly basis.

The matches resulting from using an assignment model for a particular rating, based on a week's data, show only those policy tradeoffs for that week and for that rating. The matches do not represent the impact of multiple policy objectives for a whole year or for a group of ratings. A technique is needed to incorporate the detailed weekly operational problem for each rating into an aggregate model capable of describing the overall, long-term relations among policies. By using the aggregate model, the

decision maker will know the estimated impact of the policy before a decision is made. It will help the decision maker to modify the policy or to initiate new policies that have the desired long-term effects.

### Objective

The objective of this research is to explore technological advances that make it possible to develop a long-term policy planning model that is linked to the operational model.

### General Approach

The policy model was formulated as an aggregation/disaggregation problem, considering all the complexities in linking aggregate plans to disaggregate decisions. The approach involves an integration of deterministic simulation and optimization techniques. It is characterized by multiple goals, multiple time periods, multiple levels of decision making, and a dynamic feedback structure in a large scale system.

### Results

A framework for linking operational decisions with policy planning was developed. First, a policy projection model was constructed from the disaggregated models to project the possible impact of policy options on policy goals such as PCS cost, job priority achieved, and duty preference met. Network programming and regression techniques were used. Second, a feedback control scheme was developed to steer the policy action in

the direction of achieving optimum and stable policy goals. Statistical decision rules and control limits were derived for the feedback system.

Data for the mess management (MS) specialist, storekeeper (SK), and yeoman (YN) ratings were selected to test the model. Various changes in objective functions, such as reordering of policy priority, were tested and analyzed. Aggregate regression models were constructed to measure the impact of the policy changes. Utilizing the policy impact model, a projection of the impact of planned policy changes were made and analyzed. When the projected impact shows a goal achievement level outside the control limits, another policy option is generated to obtain results more in line with the target level of goal achievement. This method will minimize the chance of overshooting and undershooting and produce a more stable result in policy goal achievement.

### Plans

The research will improve the current assignment decision process for policymaking. The work will be transitioned into an existing 6.2 project (Assignment Technology).

### Expected Benefit

The Navy spends hundreds of millions of dollars a year for advanced technical training and permanent change of station (PCS) cost. Development of a technique for policymaking will improve personnel readiness as well as resource utilization.

## Biography



**TIMOTHY LIANG** is an Operations Research Analyst in the Manpower Systems Department at NPRDC. His research specialty is large scale optimization techniques and applications to manpower problems. For several years he served as a principal investigator and project leader for developing optimization models to improve the enlisted personnel assignment process. Trained as a mathematical economist, he started his operations research work with a petroleum company at Los Angeles and worked as an economist with the State

of Hawaii before he moved to NPRDC in 1980. He is a member of the Operations Research Society of America and the Institute of Management Sciences. He has authored or co-authored many professional and technical papers. One of his journal articles was selected by NPRDC as the best paper in 1986. Another of his journal articles was selected by the Military Operation Research Society of America as the best paper in 1987.

# MODELS FOR CALIBRATING MULTIPLE-CHOICE ITEMS

James Bradford Sympson

---

*Dichotomous (right/wrong) scoring of multiple-choice test questions does not distinguish among the various wrong answers chosen by examinees. Wrong answers can supply valuable information about an examinee's capabilities. In this project, new item-response models and a polychotomous item-scoring procedure were developed. Application of this new technology to military selection, classification, and achievement testing will improve personnel decisions.*

---

## Background

### Mental Testing

Selection, classification, and training of enlisted military personnel all depend heavily on objective mental tests. Mental tests are used for selecting and classifying individuals who lack specialized training or experience and must undergo entry-level training in preparation for their military job assignments (Department of Defense, 1984). Tests are also used to assess student progress in entry-level and advanced military training courses.

The Armed Services Vocational Aptitude Battery (ASVAB) is used by the military services to select and classify civilian applicants for enlistment. Since the military services promote from within, the quality of personnel accepted for initial entry ultimately determines the quality of personnel available for the

upper enlisted ranks. Thus, both short- and long-term outcomes rely heavily on selection and classification decisions made with the help of mental tests.

In military training courses, scores on achievement tests are used, along with other information, to evaluate student mastery of course subject-matter. Mastery of the material taught in a training course usually has a strong influence on the quality of later on-the-job performance.

### Multiple-choice Questions

Mental tests often contain multiple-choice questions. Although implementation of computerized testing systems in personnel selection, classification, and training will probably reduce the number of tests administered in a paper-and-pencil format, multiple-choice questions will continue to be widely used. Even when an examinee is asked to enter a "free

response" on a computer (which requires the examinee to recall, rather than recognize, the correct answer to a question), the computer must assign the response to one of several predefined, mutually-exclusive categories. Thus, even when test questions are not presented in a multiple-choice format, they will often be scored as if they were multiple-choice questions.

### Problem

In current applications of multiple-choice questions to mental testing (e.g., in the ASVAB and in training courses), examinee responses are scored as either correct or incorrect. This dichotomous item-scoring procedure does not distinguish among the various incorrect answers that examinees select. Information about an examinee's level of knowledge that could be extracted from wrong answers is lost.

Also, currently-used item-response models fail to "fit" a portion of the multiple-choice questions that are written by test developers. If an item-response model is to be used, items that do not fit the model must be set aside. This reduces the number of items that are available for use during testing.

### Objective

The objective of this project was to develop new psychometric (psychological measurement) procedures that would extract additional information about an examinee's level of knowledge from the examinee's wrong answers to test

questions. It was anticipated that such procedures would increase the reliability of test scores, thus supporting improved personnel decisions in military selection, classification, and training.

### Progress

FY87 was the final year of funding for this project as an Independent Research (IR) effort. Following are the major accomplishments of the project:

1. Several polychotomous item-response models were developed and tried out using available test data (Sympson, 1983, 1986a, 1986b, 1987b). The most promising of these models will be used in a follow-on Independent Exploratory Development (IED) project.
2. A computer program that computes scoring weights for all the response options of a multiple-choice item was developed (Sympson, 1984). If the scoring weights derived by this program are used to score personnel tests, the reliability of those tests will increase (Sympson, 1987a).
3. A new family of statistical distribution functions and a computer program that fits this distribution function to sets of test scores was developed (Sympson & France, 1984).
4. Research results indicate that the new technology developed in this project can increase test reliability by an amount that is equivalent to a 20 percent increase in test length (Sympson, 1986b, 1987b).

**Example**

A 25-item test of quantitative reasoning ability taken by 1300 Marine Corps recruits was analyzed using the item analysis program developed in this research. One of the questions in the test was the following:

A key rack has 8 rows of hooks. Each row has 6 hooks. If 25 percent of the hooks have keys on them, how many hooks are empty?

- a. 12
- b. 16
- c. 32
- d. 36

Solving this problem requires three steps:

Step 1:  $8 \times 6 = 48$  hooks

Step 2:  $48 \times .25 = 12$  hooks with keys

Step 3:  $48 - 12 = 36$  hooks are empty

Of the Marine recruits tested, 33 percent selected option "d," as the correct answer. Another 33 percent selected option "a." The remaining 34 percent selected either "b" or "c." Apparently, individuals who selected option "a" completed the first two steps in the solution and then stopped. Although option "a" is incorrect, choosing this option clearly indicates a higher level of ability than choosing either "b" or "c," which are unrelated to the sequence of steps required to solve the problem.

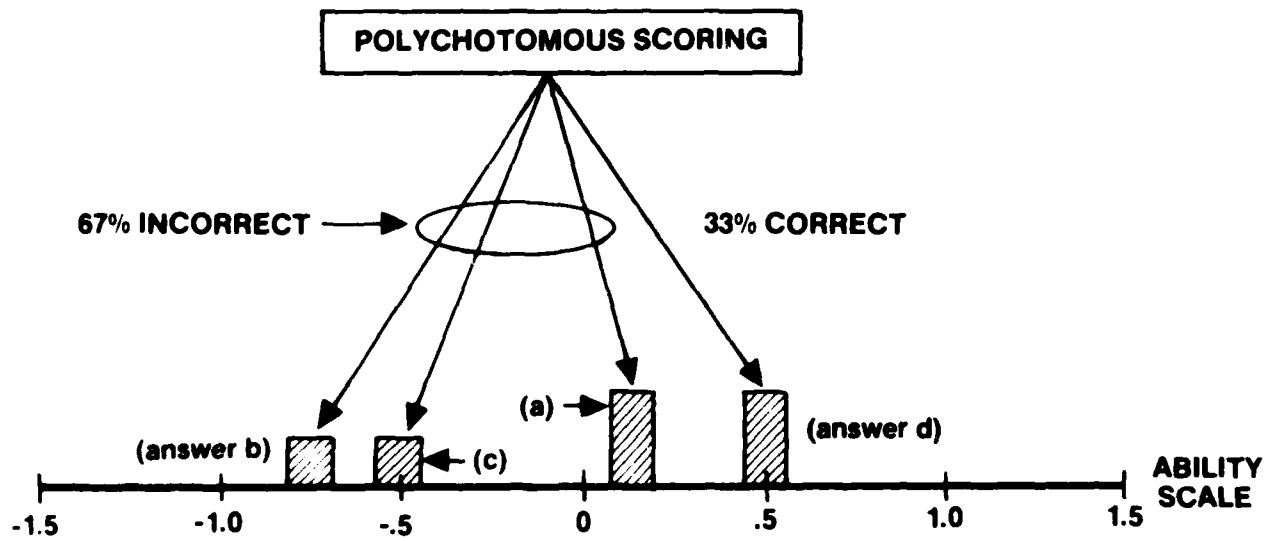
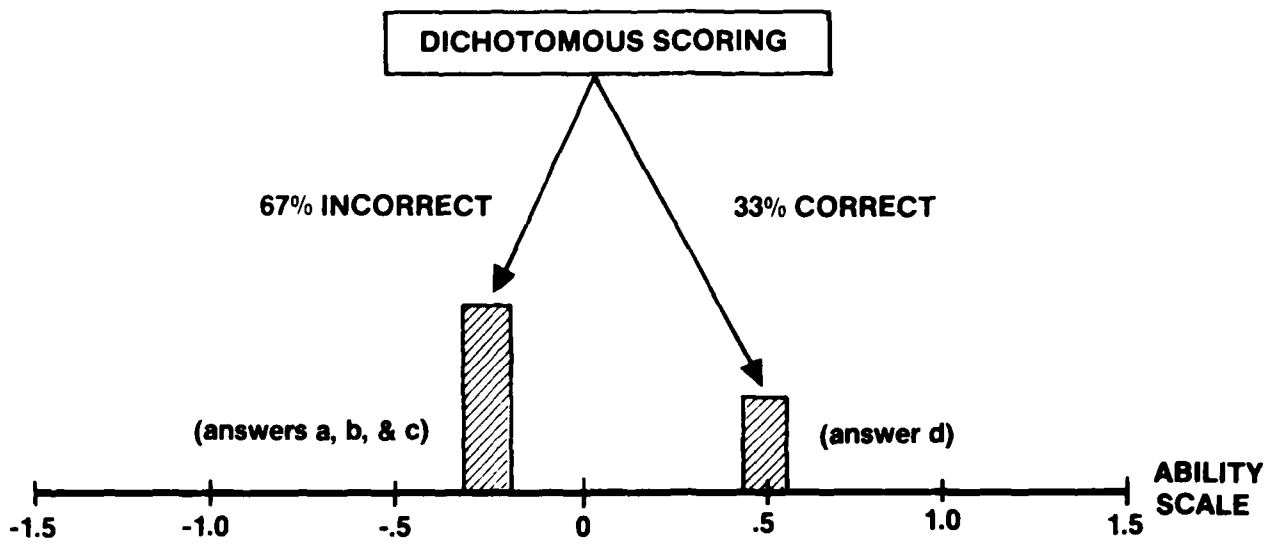
The upper portion of Figure 1 shows the result of scoring this four-choice item dichotomously. Examinees who

selected the correct answer were assigned a positive ability-level estimate, while examinees who answered incorrectly were assigned a negative ability-level estimate.

The lower portion of Figure 1 shows the result of scoring this same item polychotomously. Examinees who answered correctly were assigned the same ability estimate as before, but examinees who answered incorrectly were assigned three different ability estimates, depending on which incorrect answer they chose. In particular, examinees who selected response-option "a" received an ability estimate that is positive, but lower than the one assigned to examinees who answered correctly. Sorting people who answer incorrectly into different groups provides additional information about their mental ability and serves to increase test reliability.

This example demonstrates how additional information about an examinee's capabilities can be extracted by considering which incorrect answers have been selected. It also shows that treating all wrong answers as equivalent can be unfair to those examinees who have given partially-correct answers.

The importance of option "a" in this item was discovered using the psychometric procedures developed in this research. These procedures are based on statistical analyses of examinee item responses. They do not require one to read each question in an attempt to discover the relationship between ability and wrong answers.



Polychotomous scoring allows us to distinguish among people who answer incorrectly. This increases test reliability.

Figure 1. Models for calibrating multiple-choice items.



### Benefits

1. Empirical results (Simpson, 1986b, 1987a, 1987b) indicate that the polychotomous item scoring methods developed in this research do provide additional information about examinee ability. Application of these methods will allow us to shorten mental tests by about 20 percent, without sacrificing test reliability.

2. The best polychotomous model developed in this research has "fit" every test item to which it was applied. Thus, if this model is implemented, more of the test questions that are written can be used.

3. Our procedures allow test developers to identify test questions and response alternatives that are especially good or especially poor indicators of ability or knowledge, and aid in determining the nature of the processes that underlie examinee responses.

All of these benefits will serve to improve personnel decisions that are made in military selection, classification, and training.

### Plans

During FY88, this project will be transitioned to the Center's IED program. In the coming year, the most promising polychotomous item-response model will be applied to a wider variety of test questions. We will also document the various computer programs that have been developed and will report our research findings in the technical literature.

During FY87, collaborative research on polychotomous item-scoring procedures was initiated with Dr. Thomas Haladyna of Arizona State University. This collaboration will continue during FY88. Also during FY87, a 2-hour symposium on polychotomous item-scoring procedures was organized. This symposium will be presented at the 1988 Annual Meeting of the American Educational Research Association.

### References

Department of Defense (1984). *Test Manual for the Armed Services Vocational Aptitude Battery* (DoD 1304.12AA). North Chicago, IL: U.S. Military Entrance Processing Command.

Simpson, J. B. (1986b, August). Extracting information from wrong answers in computerized adaptive testing. In B. F. Green (Chair), *New developments in computerized adaptive testing*. Symposium conducted at the meeting of the American Psychological Association, Washington, DC.

## Biography



**J. BRADFORD (BRAD) SYMPSON** is a Personnel Research Psychologist in the Testing Systems Department at NPRDC. Brad is one of a handful of individuals in this country who have developed a high level of expertise in Item Response Theory (IRT). IRT is a new approach to mental testing that is revolutionizing the way tests are designed, administered, and scored. After completing advanced graduate work in psychometrics (the theory of mental testing) at the University of Minnesota, Brad spent 2 years at

Educational Testing Service working with Dr. Frederic M. Lord, one of the initial developers of IRT. Brad came to NPRDC in 1981 to work on the Joint Services Computerized Adaptive Testing (CAT) Project. Since coming to NPRDC, Brad has developed several new statistical procedures and IRT models for use in CAT. Brad is a member of the American Educational Research Association, the American Statistical Association, the Military Testing Association, the National Council on Measurement in Education, the Psychometric Society, the Personnel Testing Council of Southern California, and the Personnel Testing Council of San Diego. In addition to serving as a manuscript reviewer for several journals published by these professional organizations, Brad is currently the Vice-President for Research of the Personnel Testing Council of San Diego.

## BRAIN MECHANISMS FOR HUMAN COLOR VISION--IMPLICATIONS FOR DISPLAY SYSTEMS

Leonard J. Trejo  
Gregory W. Lewis

---

*The use of color in military displays is increasing, but the impact of color on the human operator is poorly understood. One important problem is the appropriate selection of color contrast for display elements. Present methods of predicting the effectiveness of color contrast in displays are based largely on behavioral threshold data, which may not be applicable to performance on dynamic visual displays. We have found that the sensitivity of individual subjects to dynamic color contrast in computer displays can be accurately assessed by visual evoked potentials (EPs). In addition, EPs are providing new insight into the mechanisms that subserve chromatic discrimination (sensitivity to color differences), which is critical for the prediction of individual performance on color-coded display systems.*

---

### Problem

The interface between human operators and complex military systems is increasingly dependent on visual information displays. With the proliferation of computers as display drivers, much more information can be presented on visual displays than the operator may effectively use. Successful design of visual displays must consider sensory, perceptual, and cognitive processes of the human operator. One focus area in visual display research is the use of color to increase the quantity and quality of information presented to the

human operator. However, the use of color in displays is proceeding without a thorough understanding of the impact of color on the human operator. In particular, most of our knowledge about human processing of color derives from behavioral research with static color displays (Burnette, 1985; Hardesty & Projector, 1973; Heglin, 1973; Meister, 1984; Merrifield & Siverstein, 1986; MIL-STD 1472C, 1981; Wagner, 1977; Wyszecki & Stiles, 1982). Little is known about the dynamics of human color processing, and even less is known about the brain mechanisms that subserve color vision.

## Background

Earlier research at NPRDC has shown that measures of brain electrical responses to sensory stimuli, known as evoked potentials (EPs), may assess unique process-related variance that relates to human performance. For example, the performance of individuals on a complex air defense radar simulation was correlated with the amplitude of visual EPs produced by a series of visual probe stimuli presented during simulation performance (Trejo, Lewis, & Blankenship, in preparation.). Other relationships between EPs and performance have been demonstrated (Lewis, 1983a, 1983b).

Other research has shown that color vision is three-dimensional and that its three dimensions are subserved by three distinct brain mechanisms (reviewed by Boynton, 1979). These include two chromatic (color-sensitive) mechanisms, red-green (R-G) and blue-yellow (B-Y), and one achromatic (A) or black-white mechanism. The activity of the chromatic mechanisms is thought to mediate chromatic discrimination, which is the ability of the visual system to discriminate colors that differ only in hue or saturation, but not in intensity (i.e., luminance).

The task of the display designer is complicated by the fact that chromatic discrimination varies across stimulus conditions. Chromatic discrimination thresholds measured under one set of spatial and temporal stimulus parameters are not necessarily valid under another set of stimulus

parameters. The designer must often rely on inappropriate data, or worse, on no data at all, in specifying color contrast for information displays. Variations also exist both between individuals and within an individual on a day-to-day basis and may reflect stress, fatigue, drug, or other biochemical effects. Even less is known about these variations than those that occur with stimulus conditions.

EP measures related to chromatic discrimination were first reported by Riggs and Sternheim (1969). Since then, little of practical significance has been made of this important finding. One possible application of this finding is the use of EPs for assessing the effectiveness of color contrast in information displays. Another possibility is the use of EPs for assessing the chromatic discrimination performance of individual human subjects. Both of these issues are addressed by the research described in this report. We find that EP measures of brain mechanisms of human color vision provide new information for personnel assessment, display systems engineering, and for understanding the basic physiology of color vision.

## Objective

The goal of this research project is to identify physiological measures of human brain activity that carry information about the activity of the chromatic mechanisms of opponent process theory, and to use these measures to improve military personnel assessment and human factors engineering.

## Approach

Procedures include recording EPs produced by stimuli generated with computerized visual displays. EPs are very small voltage signals (microvolts) recorded from electrodes placed on the scalp that represent the response of the brain to sensory input. EPs are usually extracted from larger ongoing electroencephalographic (EEG) activity by signal averaging. The stimuli are presented by the method of exchange stimulation (Estevez and Spekrijse, 1982), which involves changing the color of a stimulus dynamically (over time), while holding all other parameters (e.g. size, shape, position, and texture) constant.

## Progress

In FY 86, hardware and software were developed to present exchange stimuli and record chromatic EPs. EP data were first recorded in four laboratory personnel whose color vision was tested thoroughly using clinical behavioral vision tests (Nagel anomaloscope, American Optical HRR plates, & Farnsworth-Munsell 100 Hue Test). Subsequently, chromatic EPs were recorded from 100 military personnel, during both FY86 (Aug-Sep) and FY87 (Oct-Dec). These initial findings (Trejo & Lewis, 1987) demonstrated that EPs were sensitive to pure chromatic stimulation and that there may be individual and day-to-day variability in chromatic EPs.

In FY 87, significant progress was made in the interpretation and

analysis of chromatic EPs. The number of recordings was reduced from eight to two, and the signal-to-noise ratio of the chromatic EP was increased by approximately a factor of ten. This was accomplished by bipolar recordings of the EP local to visual cortex and digital band-pass filtering. The results in five normal subjects demonstrated a marked similarity in properties of behavioral chromatic discrimination and the chromatic EP. However, more information may be seen in the chromatic EP than in behavioral measures. Specifically, EP measures provided evidence of chromatic asymmetry in the response of the brain to the exchange of complementary colors. For example, in some subjects the exchange of green to red produced a smaller EP than the exchange of red to green in a dynamic display. Such direction-specific effects are difficult, if not impossible, to measure in dynamic displays using known behavioral methods. Evidence for another kind of brain asymmetry, known as lateral asymmetry, was also provided by the chromatic EP. For example, one subject showed much larger chromatic EPs on the right side of the head than on the left.

Results in one color deficient subject (a protanopic, or red-blind subject) demonstrated that the chromatic EP may provide diagnostic information about color deficiency. This subject showed no significant chromatic EPs in response to a red-green exchange, but showed normal EPs in response to exchanges containing blue-yellow contrast.

Several contacts with the vision research and DoD research communities were made and maintained in FY87 as a result of this project. Dr. Allen Nagy, Wright State University, is a leading researcher in the area of human color deficiency and chromatic discrimination and is co-author of an NPRDC paper to be submitted to the annual meeting of the Association for Research in Vision and Ophthalmology (Trejo, Lewis, Nagy, & White). Dr. C. White, of Boyden-White Laboratories in San Diego, is also co-author of this paper. Dr. C. Tyler, of the Smith-Kettlewell Institute for Visual Science, made a presentation at NPRDC entitled "Electrophysiological Assessment of Human Visual Function." In addition, Dr. Tyler has provided valuable scientific feedback concerning the EP methodology used in this research. Dr. L. Trejo was invited to present NPRDC research on brain mechanisms of human color vision to the Human-Computer Interaction group at the Naval Research Laboratory in September, 1987. He also had valuable interaction with researchers (Drs. Coles, Wickens, Kramer, & Gratton) of the Cognitive Psychophysiology and the Aviation Research Labs of the University of Illinois, concerning research in human-computer interaction, including chromatic EPs.

### Sample Data

#### EP Recording

EPs were recorded in trials of 5 seconds' duration during which five one-second cycles of an exchange stimulus occurred. In each cycle, the

stimulus was one color for 50 percent of the cycle time and another color during the remaining 50 percent. Three trials of each exchange stimulus were presented separated by rest periods of about 10 seconds. Thus, EPs were recorded for a total of 15 cycles of each exchange stimulus. Total recording time was about 10 minutes. Electrodes were placed on the scalp over the left and right occipital and parietal areas (O1, P3, O2, and P4, referenced to nose). Signals were amplified (20,000 times), band-pass filtered (0.1-30 Hz), digitized, and stored by a computer. Off line, bipolar potentials local to the occipital areas were derived from the digitized recordings by point-by-point subtraction of the parietal from the occipital recordings separately for each side of the head (O1-P3, O2-P4). The derived EP data for each 5-second trial (5 stimulus cycles) were digitally filtered (2.5 to 10 Hz). Then the total of 15 cycles from the three 5-second trials were averaged to form a single one-cycle (1 s) average EP for each exchange stimulus.

#### Exchange Stimuli

Each subject viewed a baseline display (D0) in which a steady white field was presented, and eight exchange stimuli, designated as follows: (D1) red versus green, (D2) bluish-red versus yellowish-green, (D3) magenta versus yellow-green, (D4) reddish-blue versus greenish-yellow, (D5) blue versus yellow, (D6) greenish-blue versus reddish-yellow, (D7) cyan versus orange, (D8) bluish-green versus yellowish-red. For each

stimulus, the first-named color corresponds to the first half of the stimulus and averaging cycle. The red-versus-green and blue-versus-yellow exchange stimuli were designed to exclusively activate either the R-G mechanism or the B-Y mechanism. The stimuli were designed to be ineffective in activating the A mechanism in color-normal subjects.

### Chromatic EPs

Figure 1 shows averaged EPs derived from the right occipital and parietal areas of one subject for each of the eight exchange stimuli and the baseline condition. Each of the waveforms is the average EP to 15 cycles of the exchange of the indicated colors. The period of the cycle is 1000 ms. A vertical bar at 500 ms indicates the mid-point between the two colors of each exchange. Amplitude scale is in units of microvolts. A regular EP waveform is observed across the different color exchanges, with a positive-going peak near 150 ms and a negative-going peak near 200 ms.

### Template Measure of EP Amplitude

To measure and compare EPs produced by different exchange stimuli, a common "yardstick" must be employed. Template-based wavelet estimation (Cohen, 1986) is a powerful method for deriving a common measure of bioelectric signals. For the data shown in Figure 1, and for five other subjects, a template was constructed by averaging the EPs

across all eight exchanges (baseline excluded). The template was then fitted to the individual EPs for each exchange by linear (least-squares) regression. The slope of the regression serves as an estimate of the relative EP amplitude for a given color exchange. In Figure 1, the template, scaled by the regression slope, is superimposed on each of the nine. Apart from minor latency variations, the template provides a good estimate of the average signal amplitude for each exchange stimulus. As expected, the regression slope for the baseline condition was not significant.

### Derived Sensitivity Measure

Most behavioral research on color discrimination concerns the estimation of sensitivity to color differences by measuring color difference thresholds. By definition, thresholds are small when sensitivity is high. Conversely, thresholds are large when sensitivity is low. In order to relate our EP amplitude measures to behavioral thresholds, we computed the reciprocal ( $1/x$ ) of the EP template regression slope for each exchange stimulus. This measure exhibits the same properties as the behavioral threshold, being large when exchange EPs amplitudes are small, and small when EP amplitudes are large. The relationship between sensitivity and this measure depends on the assumption that EP amplitudes are monotonically related to the size of color differences. Previous research supports this assumption (Riggs and Sternheim, 1969).

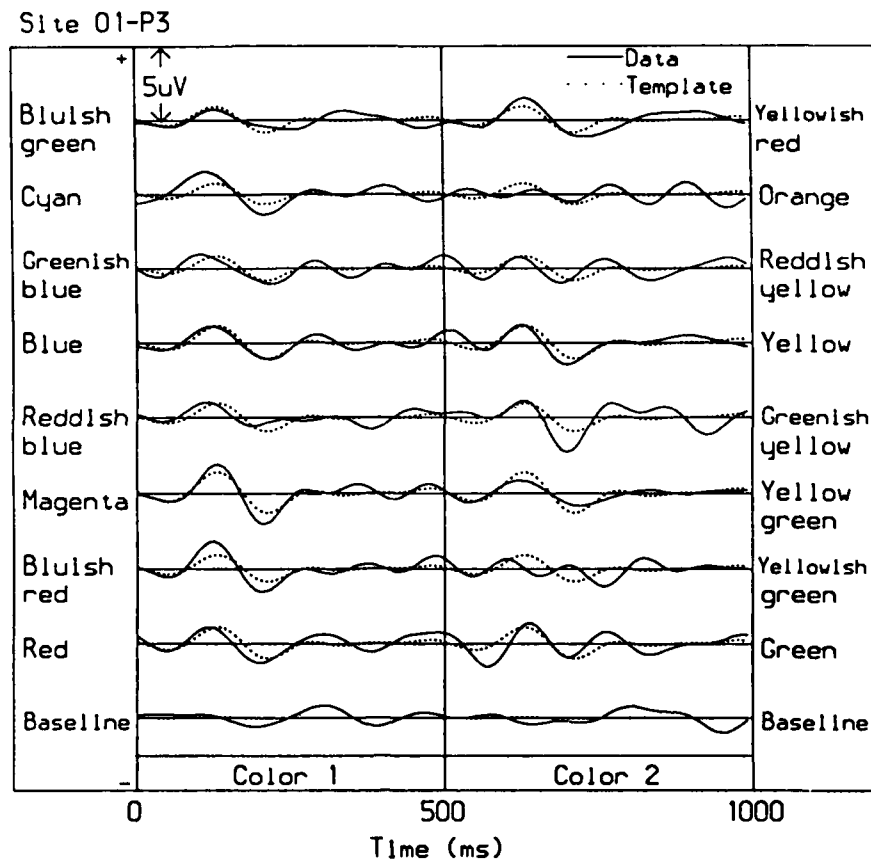


Figure 1. Averaged EPs derived from the right occipital and parietal areas of one subject for each of the eight exchange stimuli and the baseline condition.

### EP-based Sensitivity Contours

It is convenient to think of each color exchange stimulus as a "direction" in a mathematical representation of color-mixture space.

When luminance is removed from such exchanges, color-mixture space reduces to a plane. Coordinates for such a plane in terms of excitation of the R-G and B-Y mechanisms have been described as the *r*, *b* chromaticity diagram (MacLeod & Boynton, 1979). The eight exchange stimuli used in

this study were chosen to lie on vectors bisected by the coordinates of an achromatic (white) point in the chromaticity diagram, spaced at angles of  $22.5^\circ$ . When chromatic discrimination thresholds are plotted as a function of color direction with respect to a fixed point in a chromaticity diagram, the data can be fit reasonably well by an elliptical contour (MacAdam, 1942; Wyszecki & Stiles, 1982). Nagy, Eskew, and Boynton (1986) found that the best-fitting ellipse surrounding an



directions in the present study, with its major axis oriented in the direction of the exchange stimulus. This differs from the ellipse in Figure 1, which is oriented in the direction of the exchange stimulus in the color change. Because the ellipse in Figure 1 makes an angle of 45° with the major axis, the maximum color change occurs at 45°, which is not applicable to our exchange stimulus (chromatic versus achromatic).

Figure 2 shows the reciprocals of the template regression slopes of the EPs in Figure 1 plotted as a function of color-exchange direction in the r, b chromaticity diagram. The ellipse of best fit to the reciprocals of the EP template regression slope (least-squares criterion), is superimposed on the data points in Figure 2, and has its major axis oriented at 134°. This result demonstrates the high degree of correspondence to behavioral data that chromatic EP measures can obtain.

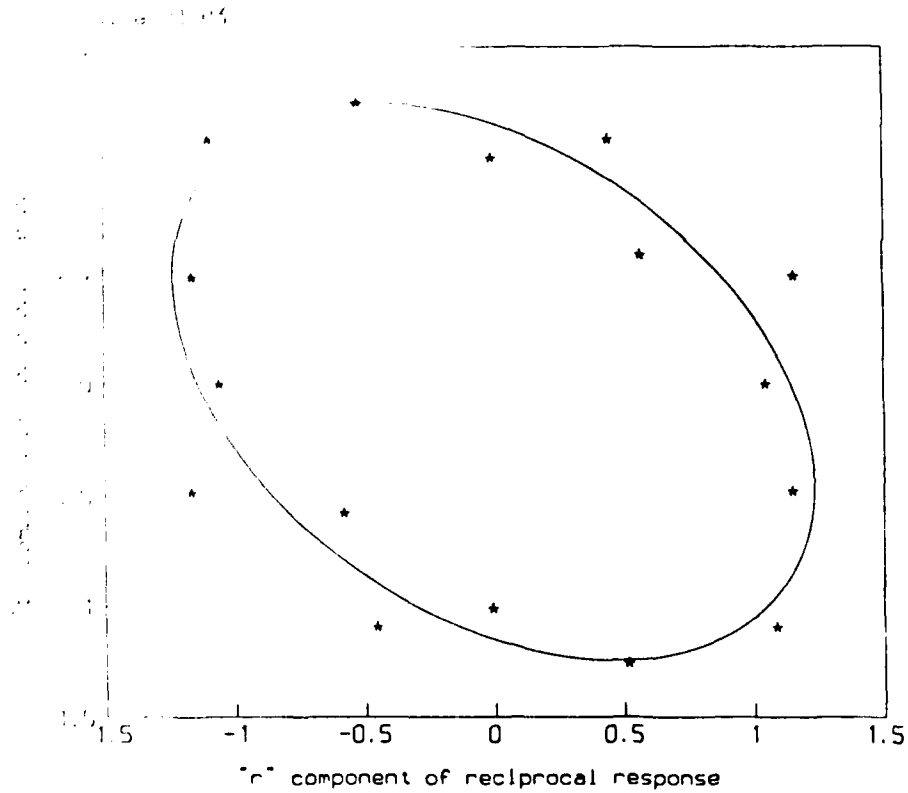


Figure 2. The reciprocals of the template regression slopes of the EPs (in Figure 1) plotted as a function of color-exchange direction in the r, b chromaticity diagram.

Trejo, L. J., Lewis, G. W., Nagy, A. L.,  
& White, C. T. (unpublished  
abstract). *Chromatic  
discrimination ellipses measured  
by visual evoked potentials.*  
Submitted to Association for  
Research in Vision and  
Ophthalmology, Annual Meeting,  
May 1988.

Wagner, D. W. (1977). *Color coding:  
An annotated bibliography* (NWC  
Tech. Pub. 5922). China Lake:  
Naval Weapons Center.

Wyszecki, G. & Stiles, W. S. (1982).  
*Color science: Concepts and  
methods, quantitative data and  
formulae.* New York: John Wiley  
& Sons.

## Biography



**LEONARD J. TREJO** was born in Mexico City, Mexico on February 24, 1955. He performed undergraduate studies at Lock Haven State College in Pennsylvania and later at the University of Oregon, receiving the B.S. (1977) degree in psychology. He performed graduate work in psychobiology at the University of Michigan and in psychology at the University of California, San Diego, with emphasis on sensation and perception. He received the M.A. (1980) and Ph.D. (1982) degrees in psychology from the University of

California for research on the neurophysiology of the pupillary light reflex and on visual sensitivity loss in hereditary retinal degeneration. From 1982 to 1984 he served as Senior Fellow in the Department of Ophthalmology at the University of Washington where he performed research on retinal toxicity, physiology of color vision, and neuroanatomy of the visual cortex. Since September 1984, he has been Personnel Research Psychologist at the Navy Personnel Research and Development Center in San Diego. His research interests include sensation and perception, color vision, visual pathways, and cognition, human performance.



**GREGORY W. LEWIS** was born, raised, and educated in the state of Washington. During his graduate work at Washington State University, he had extensive training in vision electrophysiology and neurophysiology. His doctoral dissertation was in the area of vision biometry using ophthalmic ultrasonography. From 1970 to 1974, Dr. Lewis fulfilled a military obligation as an Army officer in the U.S. Army Medical Research Laboratory, Fort Knox, Kentucky. Dr. Lewis has been with the Navy Personnel Research and Development Center

(NPRDC) since 1974. He developed and currently heads the NPRDC Neuroscience Research Project Office, Human Performance Department. This research is dedicated to developing techniques for improving the prediction of personnel performance by using neuroelectric and neuromagnetic waveform information. His areas of research interests include the psychophysiology of individual differences, digital processing of biological signals, and physiological correlates of brain and behavior.

## STATISTICAL PROCESS CONTROL AS AN ENHANCEMENT TO JOB AND ORGANIZATIONAL DESIGN

Samuel B. Landau

---

*Productivity improvement has been recognized as extremely important in the Navy's attempt to maintain readiness in light of increasingly restrictive fiscal and personnel policies. Managers are not only paying attention to technological changes but to management practices that can serve to increase the quality and productivity of products and services. Total Quality Management (TQM) is an approach that combines a set of management principles with a set of statistical process control procedures to improve product and service quality and, in turn, improve productivity. In order to implement such an approach, traditional management philosophies and practices must be changed. The ease with which these changes occur are hypothesized to be a function of the organization's present culture. Further, TQM is hypothesized to positively effect employee motivation to be more productive. The present effort identified organizational cultural factors that serve to facilitate and hinder acceptance of TQM. Measures were also obtained on baseline levels of employee motivation. Recommendations on how to facilitate the implementation of TQM were provided to the organization.*

---

### Problem

Reduced productivity growth in the United States has forced government leaders to focus efforts on improving the quality and quantity of productivity in order to ensure and maintain military security and

readiness, as well as economic survivability. Testifying before Congress (4 April 1985), Deputy Secretary of Defense, William H. Taft IV stated that "quality must (be) the keystone of our modernization efforts." The relationship between quality and productivity is being assessed

within many government organizations.

Recent government initiatives have focused on quality improvements, such as on the design and production of aircraft, ships, and weapons systems. However, little emphasis has been given to the development, implementation, and maintenance of a *management* system that will ensure that quality and productivity improvements are maintained.

### Background

An approach that attempts to improve productivity through an emphasis on improving the quality of products and/or services is TQM. TQM has been recently popularized by Deming (1982, 1985), yet has been described and applied, with some modifications, by several others (Crosby, 1979; Ishikawa, 1985; Juran, 1974). These various forms of TQM have been well received by different government agencies. The basic assertion of these orientations is that quality and productivity improvements result from a greater understanding of the processes by which work is accomplished. Corrections can be made to inappropriate work processes, thus reducing product and/or service variability. In order to make process control procedures operational, traditional management philosophies and practices need to be changed.

The ease with which these organizational changes are accepted, implemented, and maintained are hypothesized to be a function of the organization's culture, that is, the

shared values toward dimensions such as innovation, communication, participation, rewards, performance, human resource development, and customer orientation. Further, it is expected that such changes will have an effect on employee motivation. The conceptual framework being used to assess the effects of TQM on employee motivation is the Job Characteristics Model proposed by Hackman and his associates (Hackman & Lawler, 1971; Hackman & Oldham, 1976).

### Goal

The primary objectives of this effort are to determine the relationship between organizational culture and TQM, as an organizational change, and their effects on individual motivation, in terms of job characteristics.

### Approach

The sample consisted of both military and civilian employees, from top management to first level supervisors, at a naval supply organization. TQM implementation consisted of three phases. The first two phases consisted of training to top management. Phase 1 was a presentation and discussion of the philosophy and concepts of TQM, as well as of the specific activities necessary to develop long range plans. Phase 2 consisted of instruction in statistical process control procedures and in the specifics required for the implementation of TQM. Phase 3 consisted of training by top management to middle management and first level supervisors. Data was collected from questionnaires that

measured organizational cultural variables, job characteristics, and general knowledge and expectations of successful TQM implementation. Objective behavioral measures, such as absenteeism, safety (number of accidents), time to complete work orders, and error rates on filling orders were also included as part of the data collection.

The questionnaire was administered prior to Phase 1. A second administration of this questionnaire was planned after the completion of Phase 3, approximately six months after Phase 1. A second assessment of the objective measures was also planned to be taken at this time. Thus, changes in attitudes and behaviors would be determined.

### Results and Conclusions

Phases 1 and 2 were completed. Phase 3 training by top management, took the organization longer to develop than originally anticipated and had not occurred by the end of the fiscal year. Thus, the second data collection process was not able to take place over the course of the effort reported. Nevertheless, the obtained results are indicative of successful TQM implementation.

We found a general willingness to implement TQM although it was accompanied by some skepticism. This skepticism arose from a lack of information about the philosophy and application of TQM. More information was desired by middle managers and first level supervisors. While a need for more information was also

expressed by top management, they were generally more supportive of implementation. Many of the organizational culture dimensions were highly correlated with the changes being proposed by implementing TQM, such as a willingness to accept change and innovation, a focus on the customer, and addressing human resource development. Areas were also identified in which improvements could be made to facilitate the acceptance of TQM. These were basically areas of communication, such as clarifying organizational goals. Feedback of this information was presented to the organization in the form of recommendations to facilitate the process of implementing TQM.

Future efforts will be made to obtain subsequent information on the changes that may have resulted from the implementation of TQM. While the job characteristics information identified fairly good levels of motivation for individuals in the sample, the effects of the TQM implementation on them could not be assessed in the time-frame covered in this report. Collecting additional information will help to identify the relationship between TQM and individual motivation. These efforts will also indicate changes in the organizational culture dimensions and in the objective, behavioral indicators of quality and productivity.

### References

Crosby, P. B. (1979). *Quality is free*. New York: McGraw-Hill.

- Deming, W. E. (1982). *Quality, productivity, and competitive position*. Cambridge: Massachusetts Institute of Technology, Center for Advance Engineering Study.
- Deming, W. E. (1985). Transformation of western style of management. *Interfaces*, 15(3), 6-11.
- Hackman, J. R. & Lawler, E.E. (1971). Employee reactions to job characteristics. *Journal of Applied Psychology*, 55, 259-286.
- Hackman, J. R. & Oldham, G. R. (1976). Development of the job diagnostic survey. *Journal of Applied Psychology*, 60, 159-170.
- Ishikawa, K. (1985). *What is total quality control?* Englewood Cliffs, NJ: Prentice-Hall.
- Juran, J. M. (Ed.) (1974). *Quality control handbook (third edition)*. New York: McGraw-Hill.
- Taft, W. H. (1985). DoD's quality-excellence initiatives: From spares reform to streamlining. *Defense Management Journal*, Second quarter, 2-7.

## Biography



**SAMUEL B. LANDAU** is a Personnel Research Psychologist at the Navy Personnel Research and Development Center (NPRDC), San Diego, California. He received his Ph.D. at Wayne State University, Detroit, Michigan specializing in Social and Industrial/Organizational Psychology. His professional experiences have focused primarily on action research programs regarding various organizational issues. These issues have included the development, implementation, and evaluation

of ways to improve work productivity and quality in various blue and white collar organizations. Dr. Landau is presently a member of The Academy of Management and serves as a Lecturer at San Diego State University.



# THE EFFECTS OF GOALS, STANDARDS, AND REWARDS ON WORK PRODUCTIVITY AND QUALITY

D. M. Nebeker,  
B. C. Tatum,  
B. L. Cooper

---

*A productive defense establishment is vital to the Navy's ability to fulfill its worldwide defense commitments under current budgetary and manpower constraints. There have been many systems proposed that claim to motivate increased industrial productivity and quality. However, few studies have compared different systems directly or carefully examined different variations on the same system. This report discusses two studies that were conducted in a simulated organization. The first study compared two well known approaches (total quality management and goal setting) and the second study examined variations of a financial incentive system (three different performance-reward functions). In the first study, it was found that workers classified as underachievers were more productive when the production goals were set at high levels, but the overachievers were more productive when the goals were set at low levels. This finding for the overachievers is contrary to both total quality management and goal setting theories. In the second study, it was shown that, for employees with high ability, a stepped exponential performance-reward function motivated higher performance than either a smooth exponential function or a linear function. For low ability employees, this relationship did not hold. Both studies provide information that will prove valuable in helping the Navy develop techniques for increasing industrial productivity, quality, and efficiency.*

---

## Problem

We are experiencing a crisis in American industry. We are losing our

status as one of the most productive nations in the world and we are having an increasingly difficult time producing quality goods and

competing in the world marketplace. Primarily because of severe budgetary and manpower constraints associated with the expansion of fleet operations, the U.S. Navy has not been exempt from this crisis. Clearly, if the Navy is to meet its defense commitments, techniques must be developed and implemented that will increase the productivity, quality, and efficiency of Navy industrial activities. Many methods have been proposed recently for improving industrial performance. This report examines some of the methods that focus on goal setting, production standards, and performance-reward functions. The first simulation in this report investigates the effects of goals and standards on work productivity and quality from the standpoint of two very different philosophies. The second simulation explores the effects of three different performance-reward functions on improving work productivity.

### Simulation 1

#### Background

W. Edwards Deming is probably best known as the American whose philosophy and methods were largely responsible for the success of Japanese industry today (Gitlow & Gitlow, 1987, p.7). The sine qua non of Deming's philosophy is quality, and his system is known as Total Quality Management (TQM). Deming argues that improving quality through process control will improve productivity because of increased uniformity of the product; less rework and fewer mistakes; and reduced waste of manpower, machine-time, and materials. Other benefits

of improved quality, according to Deming, are lower costs, better competitive position, more jobs, and happier workers.

Deming (1982) has outlined 14 points that express his TQM methods for achieving improved quality and productivity. Two of these points state that managers should eliminate numerical goals and production standards. Deming's principle objection to goals and standards is that they are usually arbitrary and emphasize quantity rather than quality. But, even when these goals and standards are not arbitrary and do focus on quality, Deming still objects to their use because: (1) the employee often is handicapped by a process that does not provide the method and means to achieve the goal, (2) goals are often met with mistrust and resentment and the workers are often demoralized by their inability to meet the goals and standards, (3) the pressure to produce larger quantities of a product frequently leads workers to skimp on the quality, and (4) goals and standards that are too low lead to situations where employees hoard parts, slack off on their work when the quota is met, and work down to the standard (Deming, 1982; Gitlow & Gitlow, 1987).

If Deming's criticisms are valid, one wonders why management by objectives, goal setting, and industrial engineering standards have continued to be used so extensively by American business and industry. Perhaps the reason is because there is strong evidence that goals and standards do, in fact, lead to marked improvement of work productivity, quality, and

job satisfaction (e.g., Latham & Lee, 1985; Locke, Shaw, Saari & Latham, 1981; Mento, Steel & Karren, 1987). Locke and Latham (1984) are most forceful in their insistence that goal setting is a technique that works and they provide a clear program for implementing this technique.

The present research was designed to provide a clear test of these two opposing positions (TQM and goal setting) by addressing several questions: (1) Does the introduction of high production standards lead to a lowering of work quality? (2) Does the assignment of low production standards lead workers to retard their production levels (i.e., work down to the standard)? (3) Are production and quality related to individual differences among the workers (e.g., differences in achievement levels)?

### Approach

The research was conducted in the Organizational Systems Simulation Laboratory (OSSLAB) at the Navy Personnel Research and Development Center. College students were hired as "Data Base Operators" to enter and maintain a computerized data base and were paid an hourly rate of \$5.11. The employees worked 4 days a week, 4 hours a day, over a period of 2 weeks in a simulated organizational setting. The simulated organization was used to establish experimental control while at the same time allowing greater generalizability than a typical laboratory setting.

The research design was a 2 x 6 mixed factorial design. The within-subject factor was the work week

(baseline week versus treatment week). The levels of the between-subject factor consisted of two control groups and four standards groups. The control groups received no production standards during the treatment week. One of the control groups received no performance feedback and the other control group did receive feedback. The four standards groups consisted of two groups who received high production standards (110% and 120% of baseline keystroke rate) and two groups who received low standards (80% and 90% of baseline keystroke rate) during the treatment week.

Data were continuously and automatically collected by the computer workstations and included such measures as time spent on tasks, time and frequency of rest breaks, and keystrokes per hour. There were two work samples obtained from all subjects (one on the first day and one at the end of the baseline period) that served as performance-based ability measures. On the last work day, the workers were asked to complete a questionnaire that asked a wide range of job related questions. At the conclusion of the study the data entries made by the subjects were compared to a "purified" (error free) data base and the quality of their work was evaluated.

### Results and Conclusions

The data most pertinent to the focus of this report were the productivity and quality measures. Productivity was measured in terms of keystrokes per hour and quality was measured in terms of the percent of incorrect characters entered into the data base.

With respect to productivity, goal setting theory and TQM make essentially the same predictions: Those workers assigned high standards should be most productive and the workers given low standards should be least productive. The important difference between the two approaches, however, is what is claimed about work quality. With respect to quality, TQM theory predicts that the workers assigned high standards will achieve their production quotas by lowering the quality of their work. Goal setting theory, on the other hand, claims that production goals affect quality only under certain circumstances. In the context of this simulation, assigning production goals will have no effect on quality according to goal setting theory.

Figure 1 plots work productivity and quality during the treatment week for the four experimental (standards) groups. (The control groups are not shown, but it should be noted that the standards groups were more productive than the control groups.) The subjects were divided into underachievers and overachievers, based on how well they performed during the baseline period (first week) relative to their task ability. Specifically, workers who performed higher than expected, based on their work sample (ability) scores, were classified as overachievers. Likewise, workers who scored lower than expected were classified as underachievers. The data plotted in Figure 1 show that, for the quality measures, there were no substantial differences between the standards groups or between the over and

underachievers. The productivity data, on the other hand, reveal a significant interaction between levels of the standards and the degree of achievement. As Figure 1 shows, it is the underachievers who show a steady increase in performance as the level of the standards are increased. The overachievers, on the other hand, show just the opposite pattern. It is as if the overachievers get frustrated and discouraged when faced with a high production standard.

When we look at the results for the underachievers, the productivity data are consistent with both the goal setting and TQM approaches (i.e., productivity goes up with higher levels of the standard). Neither theory, however, could have anticipated the results for the overachievers. Our post hoc explanation is that the overachievers were working very hard during the baseline period, and they got discouraged and frustrated when asked to work even harder during the second week.

The results from the quality data clearly do not support TQM. Contrary to the prediction made by TQM theory, the productivity gains observed by those workers in the standards groups did not come at the expense of quality.

In general, the results provide support for both TQM and goal setting theories. The results supported the claim by both theories that productivity is directly related to the level of the standard, but only for those workers classified as underachievers. The results failed to support Deming's claim that workers given high standards would lower the quality of

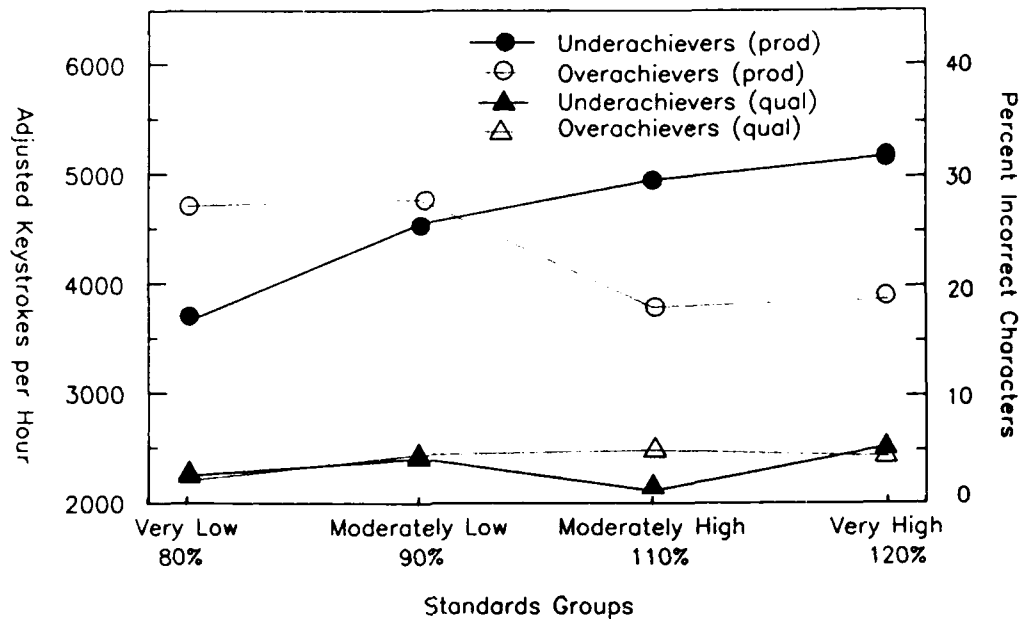


Figure 1. Work productivity and quality for the experimental (standards) groups as a function of achievement.

their work. However, if our explanation of the overachievers is correct, Deming may be justified in his claim that some workers become dissatisfied, resentful, and demoralized when faced with high production quotas. As a practical matter, these findings indicated that establishing high goals and standards may not be good for all workers. If we want to encourage the high achievers in the work force, we may be better off by not setting our production quotas at levels these workers perceive as too high.

It is important to point out that, in this first simulation, the workers were not paid any financial incentives or bonuses for working up to or above the standards. In the second simulation, we examined the effects of financial rewards for working above standard.

## Simulation 2

### Background

The use of financial incentives as a means to increase individual and group productivity has gained a resurgence in interest in recent years. This resurgence is partly a function of the recognition of the critical productivity problem we face in the U.S., and partly because recent evidence has shown financial incentives have a strong positive impact on performance (e.g., Locke, Feren, McCaleb, Shaw & Denny, 1980; Nebeker & Neuberger, 1985).

In spite of this evidence, the use of financial incentives as a means to increase productivity remains a controversial issue (Belcher, 1974; Lawler, 1981).

Preliminary research conducted in the OSSLAB has explored the impact of sharing rate, which is the amount of savings from increased productivity that is shared with the employee. An area of research closely related to sharing rate is the performance-reward function. Most financial incentive systems pay rewards as a linear function of performance. With a linear function, the employee's sharing rate is constant for all levels of performance. There is reason to question the value of this practice, however. Use of a linear function implies that the motivating value of incremental incentive increases is equal at all levels. This in turn assumes either that effort is not an important consideration in determining performance or that the relationship between effort and performance is itself linear.

The first assumption is addressed by research (Kopelman, 1977) that has suggested an individual's motivation is based on a concept called "return on effort." This concept implies that the impact of a reward offered for performance at a certain level is determined by a comparison of the sharing rate with the effort required to achieve that level of performance. Therefore, a reward amount offered may be a weak incentive for high, or difficult, performance levels, even though it is proportional to the increase in performance required to obtain it. Return on effort also implies that sharing rates may be larger than necessary to motivate individuals to improve their performance at relatively low, or easy, performance levels.

This nonlinear relationship between effort and performance results in an identical nonlinear relationship between effort and reward, because the performance-reward relationship is linear. Based upon this logic, it is hypothesized that reward systems that have an exponential performance-reward relationship will be more effective than typical linear performance-reward systems because the effort-reward relationship will approach linearity.

Further, the goal setting literature (Locke, et al., 1981) indicates that performance systems with specific, difficult, but accepted goals will increase performance over systems without goals. This conclusion suggests that a reward system with a "stepped" performance-reward function in which rewards jump to a higher level at specified intervals might represent to employees a series of difficult and specific goals that could motivate greater performance improvement than a smooth performance-reward function. Therefore, a theoretically superior variation on an exponentially accelerating function might be a function in which the rewards are stepped in an exponential fashion to represent specific and difficult goals.

Based on the above discussion, three reward systems were compared in their ability to motivate improved performance: linear, exponential, and stepped exponential. These three functions are illustrated in Figure 2. The research tested whether the use of performance-reward functions offering positively accelerating reward

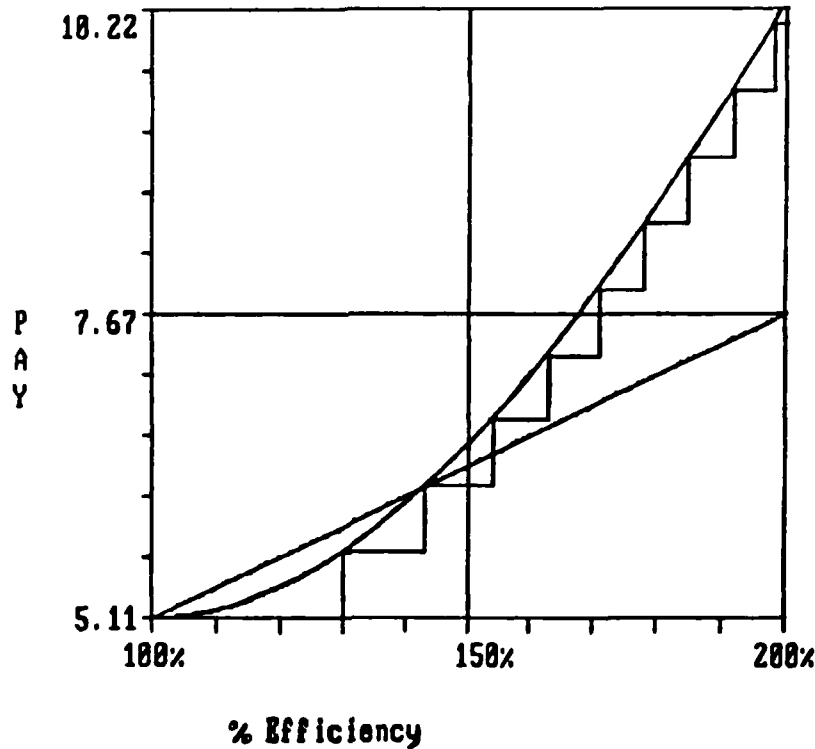


Figure 2. Linear, smooth exponential, and stepped exponential performance-reward functions.

magnitudes (exponential and stepped exponential) for equal increases in performance are more effective than linear systems.

### Approach

As in the first study, this research was conducted in the OSSLAB. This experiment investigated the effects of different performance-reward functions on employee performance. Twenty-eight proficient keyboard operators were recruited and hired (at \$5.11 per hour) to enter and maintain references in a data base designed to allow for search and retrieval of scientific literature. The employees were divided into three shifts that worked two 4-hour shifts per week for 6 weeks.

Each shift worked for 7 days in a baseline, or control, condition before the introduction of performance standards and financial incentives. Once employee performance had stabilized during the baseline condition, individual performance standards were set based on performance on a work-sample test administered on Day 5 of the baseline period. An equation derived from results in a previous OSSLAB study was used to set the performance standard.

Due to the different performance-reward functions for the three shifts, it was necessary to equalize incentive pay for the level of expected performance improvement under incentive conditions. Based on results

from a previous OSSLAB study, earnings were equalized at 43 percent of the performance standard, as shown in Figure 2.

### Results and Conclusions

The results of the experiment are shown in Figure 3, in which high and low ability performance was plotted. For employees with greater ability, performance improved more in the stepped exponential condition than in either the smooth exponential condition or the linear condition. None of the other differences were significant. This finding supports the Locke et al. (1981) assertion that difficult, specific goals, if accepted, produce greater performance improvement than easier, less specific goals. In this study, high ability employees appeared to treat each higher step of the stepped exponential system as a specific, difficult goal that motivates higher performance. By contrast, employees with less ability apparently did not accept this performance goal.

There are several possible explanations for the failure of employees with low ability to perform similarly to high ability employees. It may be that less able employees lack the confidence that they can improve sufficiently to justify the additional effort required to reach the next step. These employees may have been faced with repeated failures in attempts to achieve difficult goals, and thereby may be less inclined to take the psychological risk of failure

that accepting such a goal might involve. It is also possible that both the work sample and baseline scores represented, in addition to ability, a large component of motivation as well. In this interpretation, the employees with apparent high ability might also have been more motivated toward higher performance. Responses to questionnaires confirmed that higher performing employees were more likely to set goals.

The results indicate that:

1. Although a stepped exponential performance-reward function produced the greatest amount of performance improvement in high ability employees, it may not be the optimum method for performance in employment situations where both high and low ability employees are likely to work. Low ability employees may become demotivated or alienated by an incentive system under which they feel that they have little opportunity to earn rewards.
2. On the other hand, high ability employees may become demotivated by a linear system that provides insufficient reward to motivate performance at levels well above the standard. Under this reasoning, it may be that the smooth exponential function provides the greatest opportunity for employees of all ability levels to improve their performance. In this sense, the smooth exponential function may be perceived as the fairest of all the reward systems.



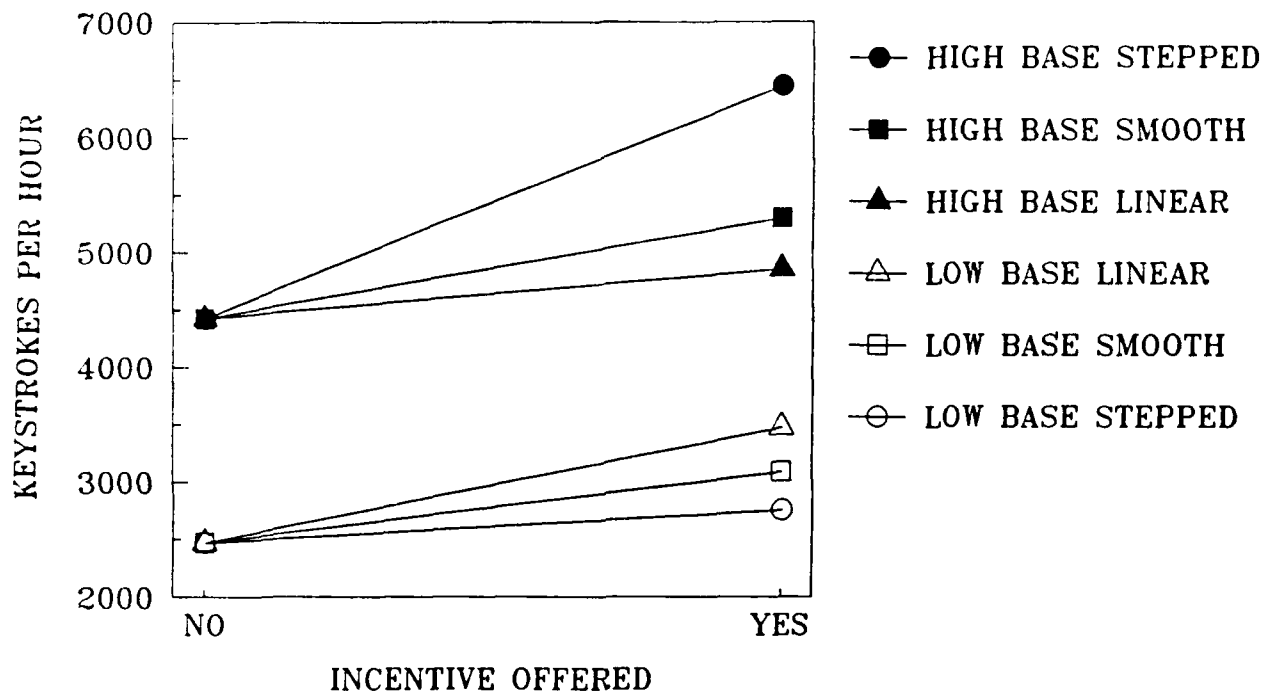


Figure 3. Performance under different performance-reward functions.

3. Each of the three designs used in this study will provide large cost savings when they are properly implemented. It is important to consider, however, not only the incentive system that offers the most cost savings, but also the perceived fairness of the system in the eyes of the employees.

### Plans

Several experiments are planned for the OSSLAB in FY88 under exploratory development funding (PE 62223). The first experiment will investigate the effects of group standards and rewards--as compared to individual standards and rewards--on productivity and work quality. Recent trends in the private sector

show an increased use of group reward systems (e.g., gain sharing plans such as the Scanlon Plan, the Rucker Plan, and Improshare). There is little evidence, however, to show if these group plans produce higher productivity and quality than individual systems, and if so, what mechanisms are responsible for the differences. The research reported here, and this future work, will be transitioned into current reimbursable field work and into the Navy Logistics Productivity Program Element (PE 63739).

### References

Belcher, D. W. (1974). *Compensation administration*. Englewood Cliffs, NJ: Prentice-Hall.

- Deming, W. E. (1982). *Quality, productivity, and competitive position*. Cambridge, MA: M.I.T. Center for Advanced Engineering Study.
- Gitlow, H. S., & Gitlow, S. J. (1987). *The Deming guide to quality and competitive position*. Englewood Cliffs, NJ: Prentice-Hall.
- Kopelman, R. E., (1977). Across-individual and within-individual return-on-effort versions of expectancy theory. *Decision Sciences*, 8, 651-662.
- Latham, G. P., & Lee, T. W. (1985). Goal setting. In E. A. Locke (Ed.), *Generalizing from laboratory to field settings: Research findings from industrial-organizational psychology, organizational behavior, and human resource management* (pp. 101-117). Lexington, MA: Lexington Books.
- Lawler, E. E., III. (1981). *Pay and organization development*. Reading, MA: Addison-Wesley.
- Locke, E. A., Feren, D. B., McCaleb, V. M., Shaw, K. N., & Denny, A. T. (1980). The relative effectiveness of four methods of motivating employee performance. In K. D. Duncan, M. M. Gruneberg, & D. Wallis (Eds.), *Changes in working life: Proceedings of the NATO international conference* (pp. 363-388). London: Wiley.
- Locke, E. A., & Latham, G. P. (1984). *Goal setting: A motivational technique that works!* Englewood Cliffs, NJ: Prentice-Hall.
- Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal setting and task performance: 1969-1980. *Psychological Bulletin*, 90, 125-152.
- Mento, A. J., Steel, R. P., & Karren, R. J. (1987). A meta-analytic study of the effects of goal setting on task performance: 1966-1984. *Organizational behavior and human decision processes*, 39, 52-83.
- Nebeker, D. M., & Neuberger, B. M. (1985). Productivity improvement in a purchasing division: The impact of a performance contingent reward system. *Evaluation and Program Planning*, 8, 121-134.

## Biography



**DR. DELBERT NEBEKER** is a Personnel Research Psychologist at the Navy Personnel Research and Development Center in San Diego, CA and a former Associate Professor of management at San Diego State University. In addition to his duties at NPRDC, he finds time to serve as an Adjunct Professor at the California School of Professional Psychology and is President of the consulting firm, Performance Matters. He is the author of numerous professional journal publications and government technical reports.

He serves on the Editorial Board for *Organizational Behavior and Human Decision Processes* and has received many awards and honors. These include the 1979 Military Psychology Award from the American Psychological Association and an Exemplary Practices in Federal Productivity award from the United States Office of Personnel Management in 1981. He is listed in *Who's Who in the West*. He received his Ph.D. in Psychology from the University of Washington and a B.S. in Sociology and Psychology at Brigham Young University.



**B. CHARLES TATUM** was born in Champaign, Illinois on October 29, 1947. He earned advanced degrees (M.A., Ph.D.) at the University of New Mexico in General Experimental Psychology. Before coming to the Navy Personnel Research and Development Center (NPRDC), Dr. Tatum was an associate professor and chairman of the psychology department at Cornell College in Iowa. While in Iowa, he was also employed as a marketing and advertising consultant and did research and post doctoral studies at the University of Iowa and the

University of Michigan. Since coming to NPRDC, Dr. Tatum has worked on several projects in the areas of organizational design and command effectiveness. He is currently involved in a project to implement "gain sharing" (an employee work-incentive program) in four Navy organizations. Dr. Tatum is affiliated with the American Psychological Association and the Academy of Management. He has published 11 papers (journal articles and technical reports) and given 16 presentations at professional meetings.

## Biography



**BARRIE COOPER** was born in New Orleans, Louisiana on July 11, 1943. He graduated in 1965 from Principia College, Elsau, Illinois, with a BA in Business Administration. After serving 10 years as an Naval Officer, Mr. Cooper worked at the Naval Air Rework Facility, North Island, in an aircraft logistics branch as a data analyst. Mr. Cooper graduated from San Diego State University in 1984 with an MBA in Management, with emphasis in Organizational Behavior. As an employee of the Navy Personnel Research and Development Center

since 1982, Mr. Cooper has worked primarily in the development of productivity and reward systems for Navy industrial activities. Mr. Cooper's other R&D interests include organizational stress and group norms. Professional contributions include first authorship on four technical reports and two papers presented at professional conferences. He has co-authored nine additional professional publications, including technical reports, technical notes, and articles published in professional journals.

## EFFECTS OF PRACTICING QUALITATIVE PROBLEMS ON TEST PERFORMANCE IN A BASIC ELECTRICITY COURSE

William E. Montague

---

*The Navy's basic electricity and electronics (BE/E) course continuously exhibits high attrition rates despite numerous changes in course content. Qualitative tests developed to diagnose student problems indicate that even the basic laws and concepts needed to maintain electronic devices are not being learned. This current evaluation studies the effects brought on by the introduction of qualitative practice problems. Student performance on course tests and in the laboratory suggest a more profound understanding of circuit functioning through the introduction of qualitative practice problems in the classroom.*

---

### Problem/Background

Student learning of "fundamental principles" of basic electricity is assumed to be essential to training for Navy electricity and electronics ratings. More than 25,000 students are required to learn these fundamentals each year. Considerable evidence has been gathered indicating that students find the material difficult to learn. Their practical skills, which presumably require the knowledge, are slow to develop. Therefore, learning difficulties need to be examined and improved instructions are needed to teach the principles and skill developments.

The current method of teaching basic electricity concepts is derived from the usual method used in

teaching physics courses. There is a dominance of teaching quantitative, formal principles first and their use in abstract problems. The course assumes that trainees have knowledge of atomic structure and electron theory, provides only a brief review, and then concentrates on mathematical formalisms and on the calculation of answers to circuit problems using Ohms' or Kirchoffs' equations. This focus has its origin in physicists' dissatisfaction with qualitative experience as the basis for theoretical understanding. As a result, they avoid teaching qualitative reasoning based on experience with phenomena and devices in favor of presenting well-structured, quantitative formalisms that avoid certain errors (Haertel, 1987). Teaching of physics and related

technology has come to be dominated by a perspective that emphasizes learning/memorizing equations and ignoring other, more qualitative understanding of devices and how they function. This approach may be generally satisfactory for handling the abstractions in mathematical notation, but it is weak in supporting the development of qualitative understanding of the concepts of electricity and functioning of simple circuits needed for practical work (Duit, Jung, & Rhöneck, 1984). The mathematical formalisms provide a major stumbling block in learning for Navy students. Academic attrition rates in the Navy Basic Electricity and Electronics course (BE/E) often exceed 20 percent, and 75 percent of that occurs during the first half of the first phase of the course.

Technicians who understand mechanical and electrical systems often reason about them at a less formal level and can adequately explain how a system or device functions and repair it without recourse to mathematical formalisms or basic principles. This competence is learned from experience with operating and maintaining machines and devices (Hegarty, Just, & Morrison, 1987). People abstract general notions about how particular devices work, and they use these ideas in attempting to understand the operation of an unfamiliar device. They use their knowledge of the components of the device and how components interact to infer the device's function. Seldom, if ever, is this knowledge described in mathematical notation. There are, obviously, different levels of understanding.

It is important to know that there are different levels of understanding of devices or machines because precise and formal levels of explanation may not be needed for some practical work. Quantitative formalisms might better be deemphasized in practical technical courses and more emphasis placed on teaching the qualitative reasoning processes of good technicians. In addition, qualitative, practical experience may provide a better foundation for understanding quantitative formalisms in courses where scientific theory is taught. Thus, practice in qualitative reasoning about device functioning may provide a useful means to promote better understanding of both device functioning for technicians, and the comprehension and use of scientific formalisms that explain them. It is the purpose of this project to explore the effect of practice in qualitative reasoning on the performance of Navy students learning basic electricity.

### Approach

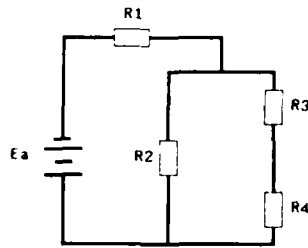
The Navy's course for teaching BE/E was analyzed and primary areas of difficulty were identified. It is an entry level training course teaching the trainee prerequisite knowledge and skills necessary for his or her follow-on job specific "A" school. The course is self-paced and consists of 52 modules that are arranged in four phases. Each phase is divided into lesson modules given to students for self-study. Module workbooks contain a lesson topic summary, programmed instruction, a narrative of the same

material, practice tests, and practice skill lessons (laboratory exercises) appropriate to that lesson. Phase one consists of 13 modules covering basic direct current (DC) and alternating current (AC) principles.

We focused on the DC portion of the first phase since most attrition occurs there. Phase one consists of an introduction to basic electricity. This includes simple electrical circuit theory, circuit symbol identification,

introduction Ohm's and Kirchhoff's electrical laws, and use of an electrical multimeter. Interviews with students and instructors and analyses of tests revealed that most of the problems are in the calculation of answers to Ohm's- and Kirchhoff's-law problems. But, students' performance solving qualitative problems, such as those shown in Figure 1, is also poor. This is important to the student because these relationships have to be understood for him to be a competent technician.

Combination Series-Parallel Circuit



**Type 1 Item:** What happens to  $E_{R1}$  if  $E_a$  decreases? **Answer** \_\_\_\_\_

That is, what happens to the voltage drop across resistor R1 if the battery voltage  $E_a$  decreases? If you think  $E_{R1}$  will increase write I or  $\uparrow$  in the space provided. Similarly, if you think that  $E_{R1}$  will decrease, write d or  $\downarrow$  in the space provided.

**Type 2 Item:** What could cause  $I_{R1}$  to increase and  $I_{R3}$  to decrease?

$E_a$	R1	R2	R3

That is, what could cause the current through resistor R1 to increase and the current through resistor R3 to decrease? In this example there is only one condition that could cause the result described. If you carefully examine the circuit and the possibilities, you will find that decreasing the value of R2 will cause  $I_{R1}$  to increase and  $I_{R3}$  to decrease. To indicate the correct answer place a  $\downarrow$  in the box below R2. If there is more than one correct answer place an arrow in the appropriate box(es).

Figure 1. Examples of two types of qualitative questions with explanations from the instructions given to students.

A method was devised to provide practice in qualitative solution of problems posed about DC circuits that instructors indicated gave students particular difficulty in learning. It took the form of a set of two types of practice problems. In Figure 1, the circuit represented is a combination series-parallel circuit. It is representative of the circuits in the last modules of the DC portion of the first phase of the course. One type of problem required a student to indicate the effects of a change in one component on the other components or circuit values. The second type of problem asks the student what change could have caused two changes observed in the circuit.

Two versions of the qualitative practice were developed. An interactive, computerized version of the practice problems was programmed in UCSD-Pascal and runs on IBM/Zenith Personal Computers. A pencil-and-paper version of the practice problems was prepared. The same problems were given in both versions.

The effects of the qualitative practice were examined by comparing test performance for three groups of students, one (21 students) receiving the computer-driven practice, another (22 students) the pencil-and-paper version. It took students about 60 minutes to complete the problems. A third "control" or comparison group (30 students) was not given the qualitative practice problems but studied the lesson module for an hour. Students were randomly selected from among those enrolled in the course during September-November 1987.

Performance on course tests were the primary measure compared. In addition, the amount of remediation required by students in the different group was measured.

### Findings

There were four measures used to compare the three groups. No statistically reliable differences were found in scores between groups on the 50 item lesson-module 6 test used in the course. On a skill (laboratory) test of 49 items, the groups receiving the qualitative practice reliably outperformed those who did not ( $F(2,69) = 6.15, p < .01$ ). Students scoring poorly on either the lesson tests or the laboratory are required to go back through the lessons and retake appropriate parts of the tests (remediation). The groups receiving qualitative practice required fewer remedial cycles than the groups receiving no such practice ( $F(2,69) = 5.22, p < .01$ ). Students who received qualitative practice outperformed the comparison group on the lesson module 7 test ( $F(2,69) = 2.993, p = .05$ ).

### Conclusions/Recommendations

Although it is not possible to draw definitive conclusions from a single empirical study, the indications are that qualitative practice can provide improvement in student learning, at least as reflected in course tests. Most important is the finding that the differences seem more robust on the laboratory tests and on the amount of remediation required. These results may indicate better student understanding of circuit functioning. Insofar as that competence is



fundamental to further technical training, it is recommended that qualitative practice be used regularly in courses that teach basic electricity. Additional research should examine the effects on learning of more substantial interventions and determine whether student ability to solve problems improves subsequent course performance.

Hegarty, M., Just, M., & Morrison, I. (1987). *Mental models of mechanical systems: Individual differences in qualitative and quantitative reasoning* (ONR 87-8). Pittsburgh, PA: Department of Psychology, Carnegie-Mellon University.

### Impact/Extentions/Transitions

Further evaluation of similar methods for training qualitative reasoning will be undertaken in the Model School Project established by CNET in 1988. A more extensive evaluation will begin in 1988 of the thesis that instructional content of basic electricity for technicians should focus on qualitative reasoning. The content should be determined by an analysis of the functional-work context rather than on the underlying scientific principles.

### References

Duit, R., Jung, W., & Rhöneck, C. v. (Eds.) (1984). *Understanding electricity: Proceedings of a workshop*. Ludwigsburg, West Germany: Pädagogische Hochschule.

Haertel, H. (1987). *A qualitative approach to electricity* (Xerox PARC Rep. No. IRL87-0001). Palo Alto, CA: Xerox Palo Alto Research Center.

## Biography



**WILLIAM MONTAGUE** is Senior Scientist in the Training Technology Department at NPRDC. His research specialty is cognition and learning. For several years he directed projects developing improvements of instructional design methods, and using computers for training. Trained as an experimental psychologist at the University of Virginia, he did research in human factors for the Navy Electronics Laboratory, taught psychology and educational psychology at the University of Illinois, and moved to NPRDC in 1972 as a project

leader. He is an active member of several professional organizations including: American Educational Research Association, Cognitive Science Society, Psychonomic Society, American Psychological Association, Human Factors Society, and Military Testing Association. He has authored or co-authored over 100 professional and technical papers, and has co-edited three books concerned with instructional psychology. He is currently a consulting editor for the *Journal of Educational Psychology*, the *Journal of Applied Psychology*, the *Human Factors Journal*, and is a peer review advisor for the Office for *Educational Research and Improvement of the Department of Education*.

**APPENDIX A**  
**PROJECT TRANSITIONS**

## Transitions

Stabilization of performance on a computer-based simulation of a complex cognitive task, to 6.2-6.3.

Policy modeling techniques for large-scale multiple objective problems, to 6.2.

Models for calibrating multiple-choice items, IR to IED and will go to 6.2 in FY89.

Effects of practicing qualitative problems on test performance in a basic electricity course, test developed used in 6.3 project.

**APPENDIX B**  
**PRESENTATIONS AND PUBLICATIONS**

## Independent Research Publications

Liang, T., & Buclatin, B. (January 1988). Improving the utilization of training resources through optimal personnel assignment in the U. S. Navy. *The European Journal of Operational Research*, V 33, (2), 183-190.

## Independent Research Presentations

- Kennington, J., Liang, T., & Thompson, T. (1987, May). *Network models and solution strategies for Navy assignment*. Joint National Meeting of the Operations Research Society of American and the Institute of Management Sciences, New Orleans, LA.
- Kidder, P. J. (In press). Participation and satisfaction in the performance assessment process. *Proceedings of the 29th Annual Conference of the Military Testing Association*. Ottawa, Ontario, Canada.
- Liang, T., & Buclatin, B. (1986, October). *Using a heuristic method to improve the efficiency of the large-scale assignment problem*. Joint National Meeting of the Operations Research Society of American and the Institute of Management Sciences, Miami, FL.
- Liang, T. (1987, May). *Monitoring assignment budgets*. Joint National Meeting of the Operations Research Society of American and the Institute of Management Sciences, New Orleans, LA.
- McMcord, M., & Liang, T. (1986, October). *Testing solution quality to the multicriteria personnel assignment problem*. Joint National Meeting of the Operations Research Society of American and the Institute of Management Sciences, Miami, FL.
- Sympson, J. B. (1983, June). *A new item response theory model for calibrating multiple-choice items*. Paper presented at the meeting of the Psychometric Society, Los Angeles, CA.
- Sympson, J. B. (1984, October). *Principal component analysis of polychotomous item responses*. Talk given at the Office of Naval Research Conference on Model-based Psychological Measurement, Princeton, NJ.
- Sympson, J. B. (1986a, April). *Some item response functions obtained in polychotomous item analysis*. Talk given at the Office of Naval Research Conference on Model-based Psychological Measurement, Gatlinburg, TN.
- Sympson, J. B. (1986b, August). *Extracting information from wrong answers in computerized adaptive testing*. In B. F. Green (Chair), *New developments in computerized adaptive testing*. Symposium conducted at the meeting of the American Psychological Association, Washington DC.

IR/IED FY87 Annual Report

Sympson, J. B. (1987a, May). *Extracting information from wrong answers in personnel testing*. Invited talk presented to members of the Personnel Testing Council of San Diego.

Sympson, J. B. (1987b, June). *Application of a polychotomous IRT model to adaptive mental testing*. Paper presented at the Office of Naval Research on Model-based Psychological Measurement, Columbia, SC.

Sympson, J. B., & France, S. L. (1984, October). *A family of distribution functions*. Paper presented at the Office of Naval Research Conference on Model-based Psychological Measurement, Princeton, NJ.



## Independent Exploratory Development Publications

- Cooper, B. L. (1987). *Stress in organizations: A review of the literature* (NPRDC Tech. Rep. 87-30). San Diego: Navy Personnel Research and Development Center.
- Cooper, B. L., Nebeker, D. M., & Riedel, J. A. (1987). *Incentive magnitude, job satisfaction, perceived stress, and performance: Interrelationships in an organizational simulation* (NPRDC Tech. Rep. 87-29). San Diego: Navy Personnel Research and Development Center.
- Nebeker, D. M. (in press). Automated monitoring, feedback, and rewards: Effects on workstation operator's performance, satisfaction, and stress. *Proceedings of the Second International Federation for Information Processing Conference on Human Computer Interaction*. Stuttgart: F. R. Germany.
- Riedel, J. A., Nebeker, D. M., & Cooper, B. L. (1987). *Incentive magnitude, job satisfaction, perceived stress, and performance: Interrelationships in an organizational simulation* (NPRDC Tech. Rep. 87-15). San Diego: Navy Personnel Research and Development Center.
- Riedel, J. A., Nebeker, D. M., & Cooper, B. L. (in press). The influence of monetary incentives on goal choice, goal commitment, and task performance. *Organizational Behavior and Human Decision Processes*.
- Tatum, B. C., Nebeker, D. M. (1987). A model of work strategies (in preparation). To be submitted to *Human Performance*.
- Tatum, B. C., & Nebeker, D. M., Cooper, B. L., & Riedel, J. A. (1986). *Work strategies: The development and testing of a model* (HFOSL Tech. Note 72-86-5). San Diego: Navy Personnel Research and Development Center.

## Independent Exploratory Development Presentations

- Cooper, B. L., & Nebeker, D. M. (1986, August). *Incentives, performance, and perceived stress in an organizational simulation*. Paper presented at the annual meeting of the American Psychological Association, Washington, D.C.
- Nebeker, D. M. (1985, April). *Employee gainsharing for office professionals and productivity: Are we missing an opportunity?* Paper presented at a press conference sponsored by Professional Secretaries International and Minolta Corporation, New York, NY.
- Nebeker, D. M. (1985, October). *The effects of reward magnitude and difficulty of performance standards upon individual productivity*. Paper presented at the annual meeting of the Military Testing Association, San Diego, CA.
- Riedel, J. A. (1985, August). *Some problems and rewards of doing research in an in-house R&D institution*. Invited address at the Doctoral Consortium at the meeting of the Academy of Management, San Diego, CA.
- Sorensen, S. W., & Nebeker, D. M. (1985, April). *Workload standards and the quality and quantity of productivity: Experiments in a simulated organization*. Paper presented at the annual meeting of the Operations Research Society of America/The Institute of Management Science (ORSA/TIMS), Los Angeles, CA.
- Tatum, B. C., & Nebeker, D. M. (1986, August). *A model of work strategies*. Paper presented at the annual meeting of the American Psychological Association, Washington, DC.

**APPENDIX C**  
**AWARDS AND HONORS**

**Independent Research/  
Independent Exploratory Development  
Researchers Awards and Honors**

Liang, T. Professional Publication Award, Navy Personnel Research and Development Center. (FY 1986).

Liang, T. Koopman Prize, The Operations Research Society of America (1987).

## DISTRIBUTION LIST

Assistant Secretary of Defense  
Deputy Under Secretary of Defense for Research and Engineering (Research and  
Advanced Technology)  
Assistant for Training and Personnel Technology (OUSD) (A)/R&T (E&LS)  
Assistant Secretary of the Navy (Manpower and Reserve Affairs) (OASN)  
(M&RA)  
Deputy Assistant Secretary of the Navy (Manpower)  
Assistant for MPT Research and Development and Studies (OP-01B7)  
Assistant for Manpower, Personnel, and Training (OP-987H)  
Commander, Naval Sea Systems Command  
Commander, Naval Air Systems Command  
Commander, David W. Taylor Naval Ship Research and Development Center  
Commanding Officer, Naval Coastal Systems Center  
Commanding Officer, Naval Underwater Systems Center  
Commander, Naval Ocean Systems Center  
Commander, Naval Air Development Center  
Officer in Charge, Naval Surface Warfare Center Detachment, White Oak  
Laboratory  
Director of Navy Laboratories Office (SPAWAR 05T), (SPAWAR 51E1)  
Commander, Naval Weapons Center (Code 38)  
Life Sciences Directorate (OCNR-114), (OCNR-30T), (OCNR-0722)  
Office of Naval Research, Detachment Pasadena  
Office of Naval Research, Detachment London  
Chief of Naval Education and Training (Code 00A)  
Chief, Bureau of Medicine and Surgery (MED 25)  
Commanding Officer, Naval Aerospace Medical Research Laboratory  
Commandant of the Marine Corps (MPI-20)  
Commandant of the Marine Corps (RD)  
Commander, U.S. ARI, Behavioral and Social Sciences (PERI-POT-I)  
Commander, Air Force Human Resources Laboratory, Brooks Air Force  
Base, TX  
Program Manager, Life Sciences Directorate, Bolling Air Force Base  
(AFOSR/NRL)  
Commander, Air Force Human Resources Laboratory, Wright-Patterson Air  
Force Base (AFHRL/LRS-TDC)  
Commander, Air Force Human Resources Laboratory, Lowry Air Force Base  
(AFHRL/ID)  
Director of Training, Office of Civilian Personnel Management  
Superintendent, Naval Postgraduate School  
Director of Research, U.S. Naval Academy  
Executive Director, U.S. Naval Institute  
Institute for Defense Analyses, Science and Technology Division  
Center for Naval Analyses, Acquisition Unit  
Defense Technical Information Center (DTIC)

END

DATE

FILMED

7-88

Dtic