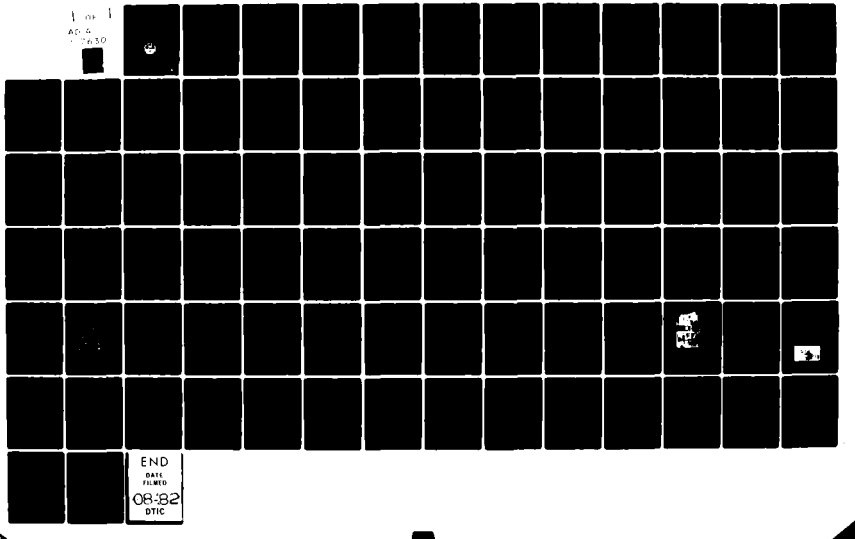
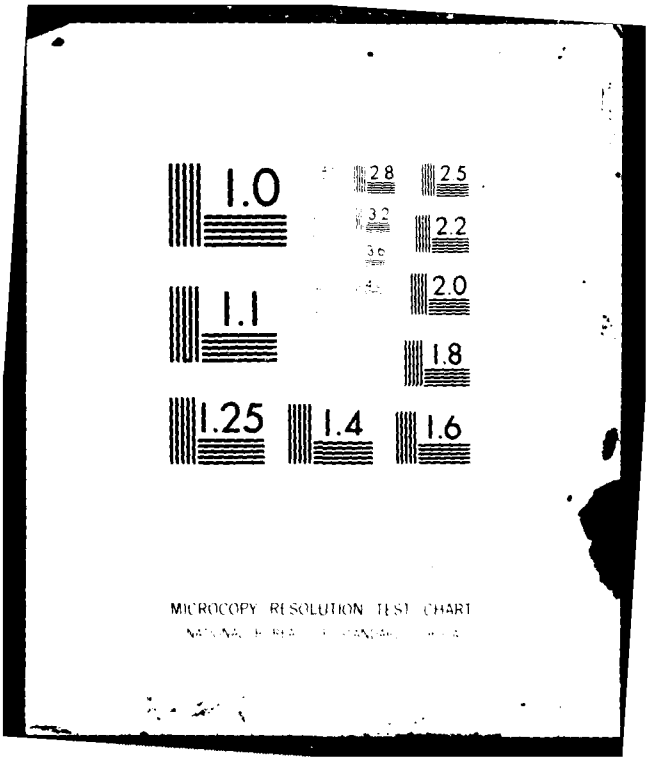
AD-A117 630 NAVY PERSONNEL RESEARCH AND DEVELOPMENT CENTER SAN D--ETC F/G 5/9
INDEPENDENT RESEARCH AND INDEPENDENT EXPLORATORY DEVELOPMENT AT--ETC(U)
JUN 82 B RIMLAND
UNCLASSIFIED NPRDC-SR-82-27 NL



END
DATE FILMED
08-82
DTIC


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6
MICROCOPY RESOLUTION TEST CHART

(12)

NPRDC SR 82-27

JUNE 1982

INDEPENDENT RESEARCH AND INDEPENDENT EXPLORATORY DEVELOPMENT AT THE NAVY PERSONNEL RESEARCH AND DEVELOPMENT CENTER--FY81

NAVY PERSONNEL RESEARCH
AND
DEVELOPMENT CENTER
San Diego, California 92152

NPRDC Special Report 82-27 June 1982

INDEPENDENT RESEARCH AND INDEPENDENT EXPLORATORY DEVELOPMENT AT THE NAVY PERSONNEL RESEARCH AND DEVELOPMENT CENTER—FY81

Bernard Rimland

Reviewed by
Richard C. Sorenson

Released by
James F. Kelly, Jr.
Commanding Officer

Navy Personnel Research and Development Center
San Diego, California 92152

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NPRDC SR 82-27	2. GOVT ACCESSION NO. AD-A117630	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) INDEPENDENT RESEARCH AND INDEPENDENT EXPLORATORY DEVELOPMENT AT THE NAVY PERSONNEL RESEARCH AND DEVELOPMENT CENTER--FY81		5. TYPE OF REPORT & PERIOD COVERED Special Report FY81 6. PERFORMING ORG. REPORT NUMBER 03-82-6
7. AUTHOR(s) Bernard Rimland		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Navy Personnel Research and Development Center San Diego, California 92152		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS PE62766N PE61152N
11. CONTROLLING OFFICE NAME AND ADDRESS Navy Personnel Research and Development Center San Diego, California 92152		12. REPORT DATE June 1982 13. NUMBER OF PAGES 82
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Independent Researach (IR)
Independent Exploratory Development (IED)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

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

This report provides synopses of FY81 IR/IED projects, the IR/IED funding profile, and a list of publications and presentations on IR/IED projects. It also includes extended articles on IR or IED efforts that have been developed over several years.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

FOREWORD

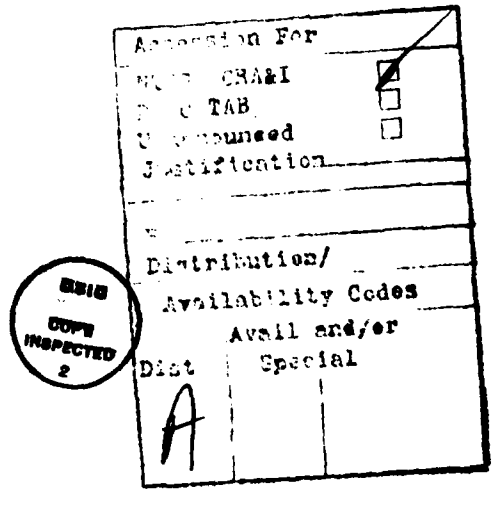
The Independent Research program at the Navy Personnel Research and Development Center has been active since the Center was formed in 1973. It is funded under PE61152N. The Independent Exploratory Development program was initiated in FY76 and is funded under PE62766N.

This report is submitted to fulfill the requirement (NAVMATINST 3920.B) for an annual IR/IED report. It provides synopses of FY81 IR/IED projects, the IR/IED funding profile, and a list of publications and presentations on IR/IED projects.

An innovation in this year's report is the inclusion of five extended articles that describe IR or IED efforts that have been developed over several years, and that may be of interest to a relatively wide readership. These articles, which appear as Appendices A through E, will be submitted for possible publication in appropriate periodicals.

JAMES F. KELLY, JR.
Commanding Officer

JAMES J. REGAN
Technical Director



CONTENTS

INDEPENDENT RESEARCH (IR) PROGRAMS

The Navy Personnel Research and Development Center (NAVPERSRANDCEN) has been conducting Independent Research (IR) since its establishment in 1973. The resources provided by this program have been used to develop research methods and techniques related to training and utilization of people in the Navy. In FYs 78 to 82, work has been concentrated in the seven areas indicated in Table 1.

Table 1

IR Funding at NAVPERSRANDCEN

	Area	Funding in Thousands of Dollars[a] FY78	FY79	FY80	FY81	FY82
ZR000-01-042						
-04.01	Measurement Theory and Methods in Personnel Research	85	60	76	83	68
-04.02	Measures of Average Evoked Cortical Potentials	55	48	57	0	0
-04.03	Retention of Quality Personnel	0	0	0	95	0
-06.01	Instructional Psychology	55	85	97	77	127
-06.02	Reading Failures	0	67	0	0	0
-08.02	Comparison of Public and Private Sector Managerial and Executive Jobs	0	0	10	0	0
-08.03	Group Performance Effectiveness	0	0	0	0	65
		195	260	240	255	260

[a]As of 1 April 1982.

The IR work that has been accomplished in FY81 is described in the following pages. In addition, Appendices A through C provide extensive articles on efforts receiving substantial IR support over 2 or more years. These articles, which describe research on psychological scale values, learning from text, and information processing respectively, were prepared for possible publication in appropriate periodicals.

MEASUREMENT THEORY AND METHODS IN PERSONNEL RESEARCH

A Procedure to Revise Estimates of Psychological Scale Values

Background

Intelligent decisions to change or terminate policies can be made only after giving due consideration to (1) the potential effects of such changes, (2) the conditions under which various effects may occur, and (3) available information about prevailing operational conditions. This project was concerned with the development and evaluation of methods that may be used by policy makers to alter estimates by taking into account both recently acquired data elements and data elements that are inconsistent with one another.

Approach

The project involved the formulation of the mathematical model, the development of specifications for use of the model, and an evaluation of the effects of adding updated information to the decision simulation process. A Navy personnel assignment model was used to investigate the proposed procedure.

Findings

The technique developed and evaluated proved useful in the application in which it was tried--the revision of a procedure used to match recruit applicants with entry level Navy jobs. The mathematical models developed at NAVPERSRANDCEN were used to integrate new data into a preexisting data base. The resulting index is now used on a daily basis, involving the classification of nearly 100,000 applicants per year.

Note. This work was accomplished under FY78-82 work unit ZR000-01-042-04-01.03: Differential Worth for Optimal Individual Placement. A more extensive discussion of the work unit is provided in Appendix A.

Student and Instructor Assessments of Computer-managed Training

Background

The Navy is experiencing severe training problems as a result of the high attrition of skilled petty officers and their replacement with new personnel, who often have difficulty in learning the high technology skills required by the modern Navy. Further, the attrition of capable instructors necessitates the increased use of extensive individualized instruction and computer-managed instruction (CMI) techniques. Although it has been demonstrated that such techniques have numerous benefits in the acquisition of new skills, many training managers at all levels feel that students and instructors have negative attitudes toward these new systems. Investigation of this problem is imperative, since attitudes can adversely affect the way in which students approach their task, the competencies they build, and the rate at which they complete training. Instructors' attitudes may have a great impact on their students. Such attitudes can be an impediment to the adoption of a potentially successful program.

An accurate assessment of attitudes of Navy instructors and trainees toward the current CMI system could lead to the identification and alleviation of significant problems resulting from the use of CMI in Navy training.

Approach

Questionnaires were developed and administered to 100 instructors and 255 trainees from five schools taught under the CMI system. Trainee and instructor questionnaires contained items exploring attitudes toward the CMI system in the learning environment, demographics, interactions with instructors or students, and motivation.

Findings

1. Trainees' attitudes toward the CMI system in the learning environment were generally favorable, while those of instructors were generally unfavorable.

2. Trainees' length of service in the Navy appears to be related to attitudes toward the CMI system in the learning environment. The longer trainees have been in the service, the more negative they tend to be toward the system.

Recommendation

The Chief of Naval Technical Training should further investigate attitudes toward CMI to identify more precisely those characteristics of CMI that lead to negative attitudes. The CMI system itself could be used to collect and analyze the attitudinal data.

Note. This work was accomplished under FY80-81 work unit ZR000-01-042-04-01.04: Unobtrusive Measures of Attitude. A more extensive discussion of this work is provided in NPRDC TR 82-19 (Robinson, Tomblin, & Houston, 1981) (AD-A109 664).

Methods for Clustering Tasks

Background

Cluster analysis is frequently used to group together occupations or jobs that are similar in terms of tasks, attributes, or other work-related requirements. Such cluster profiles are useful for aligning training curricula with actual work performed and streamlining occupational classification systems by combining administratively separate jobs into one job type. Also, there is a growing need for such job analysis methods for use in developing personnel tests, procedures, and policies that comply with federal employment guidelines. The Comprehensive Occupational Data Analysis Program (CODAP) was developed by the Air Force to fill these needs, and is in wide use by both military and civilian organizations. The data sets analyzed are typically job analysts' or job incumbents' responses to items within a structured task inventory. An important decision, one that can affect the cluster analysis solution, is the selection of a proximity measure to assess the degree of similarity among patterns of responses.

All U.S. military services use the powerful computer-based clustering procedures available in CODAP. While recent research has suggested that simple binary measures may be able to capture as much profile information as do continuous measures, there has been no empirical comparison of cluster solutions produced by the application of various proximity measures to occupational data. The present research is directed toward making such a comparison. Selection of a binary proximity measure for programming into CODAP System 80, an enhanced IBM version of CODAP currently being developed by the Department of Defense, will be based on the conclusions of this research.

Approach

Data collected by the Navy Occupational Development and Analysis Center consisted of three samples of job incumbent profiles. Each sample was comprised of 250 profiles, indicating time spent on various tasks and based on a six-point scale. Sixteen proximity matrices were derived for each sample, with each matrix based on the application of one of three continuous or 13 binary proximity measures. The CODAP clustering procedure, an average linkage procedure, was applied to seven selected proximity matrices from each of the three samples. The evaluation of the binary measures was based on the extent that the binary matrices and cluster solutions were objectively similar to those based on continuous measures.

Conclusions

1. The Jaccard and Dice proximity measures are consistently powerful measures, capable of capturing more profile information than many other binary measures.

2. The performance of a proximity measure in cluster analysis can be strongly affected by the proportion of zeros in the data analyzed.

3. The use of selected binary proximity measures will yield cluster solutions highly similar to those based on continuous proximity measures.

4. As the high proportion of zeros in the incumbent profiles analyzed are typical for this type of data set, the findings appear generalizeable to data collected from other occupational inventories.

Note. This work was accomplished under FY81 work unit ZR000-01-042-04-01.05: Methods for Clustering Tasks. A more extensive discussion of this work is provided in NPRDC TR 82-36 (Pass & Chatfield) (AD-A112 930).

Models for Personnel Management

Background

The management of human resources as a policy and planning enterprise is a relatively new phenomenon. The purpose of this effort was to assess the state of the art in human resource management for both public and private personnel systems, with emphasis on the strategic aspects of manpower planning.

Approach

The small body of pertinent scientific literature was reviewed and a managerial perspective was derived from the theoretical and technical features of manpower planning research and practice.

Accomplishments

A case study was developed that centers on the problems of acquiring and retaining engineering manpower. The problems are incurred when a high technology firm is acquired by a larger, more conventional manufacturing concern. This case study illustrates many of the interrelated factors involved in managing the internal labor market of an organization.

Along with the case study, the internal personnel structure and flow properties of public and private manpower systems were examined. After a description of internal supply and demand, consideration was given to (1) the rationale for personnel planning, (2) the relationship of a personnel system to its external environment, (3) the state of the personnel inventory, (4) methods for forecasting personnel losses and determining gains, and (5) methods for controlling the personnel system.

The case study and the assessment of personnel systems in terms of internal supply and demand represent two chapters drafted as part of a book being prepared on manpower planning.

Note. This work was accomplished under FY81-82 work unit ZR000-01-042-04-01.06: Models for Personnel Management.

A New Model for Calibrating Multiple-choice Items

Background

In the last 10 years, a new thrust in psychometric theory has emerged that may lead to a variety of important new advances in personnel testing. This approach to test construction, scoring, and evaluation is known as item response theory (IRT). The basic assumption underlying IRT is that the probability of selecting a particular response to a test question is related to a person's level of ability (or other relevant personal characteristic) by a specified mathematical function. What differentiates various implementations of IRT is the form of the mathematical function that is assumed.

Prior IRT models have two drawbacks. First, they classify responses to aptitude questions as either correct or incorrect. They make no distinction among the different incorrect responses a person might select. Information about a person's level of ability that could be extracted by considering which particular incorrect responses have been selected is lost when these models are used. Second, these models assume that the probability of a correct response to an item is a strictly increasing monotonic function of ability. In this research, evidence was uncovered that suggests this assumption is too restrictive.

Approach

Data from several widely used tests were examined. For many of the items, the proportion of correct responses failed to increase monotonically as ability increased.

If a significant number of items in a given aptitude domain have nonmonotone regressions, current IRT models cannot provide an adequate basis for personnel testing in that domain. To deal with this problem, a new IRT model based on the multivariate logistic function was developed. This model incorporates parameters for each item response. The various incorrect responses to a multiple-choice question are kept distinct and model parameters reflect the tendency for people at different ability levels to find different responses differentially attractive. Also, the model is flexible enough to allow fitting of nonmonotonic item-ability regressions when they occur.

Conclusions

If this new IRT model is adopted for the calibration of test items, two immediate benefits can be realized. First, fewer items will have to be rejected for lack of fit to the model, thus reducing the cost of test development. Second, once tests have been assembled, the use of information about which incorrect alternatives an individual has selected will improve our estimates of ability. This will increase the reliability of tests of any given length and will also allow the use of somewhat shorter tests to achieve prespecified levels of reliability. The latter possibility could lead to significant savings in testing time.

Note. This work was accomplished under FY81 work unit ZR000-01-042-04-01.07: Models for Calibrating Multiple-choice Items.

RETENTION OF QUALITY PERSONNEL

Methods for Projecting Petty Officer Retention

Background

Retention of experienced enlisted personnel during recent years has been far below the level needed for top naval efficiency, effectiveness, and readiness. The purposes of this effort were to: (1) relate the major factor in eligibility for reenlistment (i.e., recommendation by the member's commanding officer) to a set of wide-ranging experiences and assignments, and (2) conduct a search for the important correlates of reenlistment among these same experiences and assignments.

Approach

Historical data from three enlisted ratings (boiler technician, hull technician, and operations specialist) were obtained for all members processed for reenlistment/discharge during 1977-79. Recommendation for reenlistment of each member and the actual reenlistment/discharge outcome were compared with a wide range of assignment factors (e.g., duration of assignment, sea vs. shore designation, and ship characteristics). The members were subgrouped for maximal differential prediction of recommendation and reenlistment.

Findings

1. The duration of the first duty station assignment was found to be related to both recommendation and reenlistment, which occurs years later, in all three ratings.

2. The sea vs. shore designation of the first duty assignment showed virtually no relationship with either recommendation or reenlistment, while the designation of the last duty station revealed that sea duty was positively related to reenlistment.

3. The number of unauthorized absences and the type of ship for the first duty assignment interacted to predict maximally recommendation for reenlistment in all three of the ratings.

4. The following three variables interacted to predict maximally reenlistment: duration of first duty assignment, sea duty as the last duty assignment, and being married at the time of reenlistment or discharge.

Conclusions

The duration of the first duty assignment is an important long-range predictor of reenlistment, whereas sea duty is not. Short-term predictors of the last duty assignment include sea duty and being married--both lead to a greater percentage reenlisting.

Recommendation for reenlistment can be maximally predicted through the use of the incidence of unauthorized absences and ship classifications. Duration of the first duty assignment is also correlated with this variable.

Note. This work was accomplished under FY81 work unit ZR000-01-042-04-03.01: Methods for Projecting Petty Officer Retention.

Reenlistment Decision Processes

Background

There is currently a shortfall of approximately 22,000 petty officers in the Navy. Of special concern is the high rate of loss of experienced petty officers with 8 to 12 years of technical experience and training. Better understanding of the causes of poor retention is essential in formulating ameliorative policy changes. A valuable data source for identifying the causative factors is the Enlisted Separation Questionnaire (ESQ), which is completed by every enlistee who leaves the Navy. The questionnaire results have been analyzed to identify the reasons for leaving the Navy most frequently cited by the overall population. However, since an individual may cite several reasons, all with equal value, it is necessary to determine the relative importance of an individual's reasons for separation and whether or not there are groups of reasons that represent larger clusters.

The objectives of this research effort were to (1) apply innovative quantitative methods to the analysis of the ESQ to determine the relative importance of each separation reason, (2) determine "reason clusters" through data reduction methods, and (3) identify specific reason clusters that can best predict the separation behavior of subpopulations within the Navy.

Approach

In 1980, approximately 7400 regular Navy enlisted persons completed the ESQ, with approximately 72 percent leaving during or at the end of their first enlistment. The ESQ consists of two parts: several background questions, and 30 items listing reasons for leaving the Navy (e.g., "*working hours are too long*" and "too much sea duty"). Individuals responded to each reason using a 5-point scale ranging from "no importance" to "extremely important." While a small sample of people rated every item as important, 75 percent of the sample marked 10 or fewer reasons as important. The important reasons were analyzed by factor analysis, thus reducing them in number and clustering them into coherent groups. The obtained factor scores for individuals were then used as predictor variables in a discriminant analysis, distinguishing between a number of subsamples. These included technical versus nontechnical specialities, premature attrition versus completion of EAOS, critical versus noncritical ratings, etc., for each of up to four enlistments.

Findings

While pay was the reason most frequently cited (over 50% of the respondents rated pay as extremely important), over 40 percent did not rate pay as extremely important. Of special interest to this study is the analysis of reasons given by those who did not cite pay.

The factor analysis revealed eight independent factors. These factors, which were stable across different enlistments, related to problems in (1) supervision, (2) utilization of skills, (3) quality of benefits, (4) loss of benefits, (5) military regimentation, (6) nonpermanent home, (7) work conditions, and (8) pay.

Factor scores used in a discriminant analysis showed that several factors discriminated between groups. For example, the Navy's failure to use skills appropriately was a more important reason for separation for nontechnical persons than for technical persons. Poor skill utilization was said to be particularly important for those who prematurely attrited during the first enlistment.

While pay continued to be an important factor throughout the enlistments, it was not as significant for technical persons as for nontechnical persons. Pay was often cited as particularly important to those who prematurely attrited during the first enlistment.

Factors other than pay became increasingly important after the first enlistment. Specifically, lack of permanent homes and loss of benefits increased in importance, moving from the first to the fourth enlistments. The gain in importance of these reasons is not surprising, considering the increase in the number of married persons.

When critical versus noncritical ratings were compared, there were no appreciable differences in the importance of each of the factors.

Conclusions and Recommendations

The reasons given for separating from the Navy vary significantly for the different enlistment periods. The ESQ data base is a rich source of information for understanding reasons for separation. Since the factor analyses generated stable, meaningful factors, the data are well suited to further analyses.

Recommendations include further analyses of the additional separation data collected to date, and the application of multivariate techniques to scale the results on the importance dimension. Recommendations for policy changes that may emerge from these analyses include improving the quality of supervision for first-term enlisted personnel and better utilization of their skills, particularly those of nontechnical personnel.

Note. This work was accomplished under FY81 work unit ZR000-01-042-04-03.02: The Reenlistment Decision Process.

INSTRUCTIONAL PSYCHOLOGY

Enhancement of Information Acquisition and Storage

Background

A considerable amount of civilian and military education and training is based on written instructional materials, and there is a substantial body of research on factors that affect learning from text. However, much remains to be learned.

Approach

NAVPERSRANDCEN has had an ongoing program of research in this area since 1976. A series of experiments has been conducted to provide guidance for those preparing training text materials. Some of the areas investigated include (1) the effects of various types of practice questions and instructions, (2) the relation between practice and test questions, and (3) the way practice questions are presented.

Findings

1. Giving students instructions about the final test is often as effective as, and in some cases more effective than, giving adjunct questions.

2. A significant decrement in test performance occurs when lesson materials, including practice items, are not consistent with one another.

3. Student performance improves along with the percentage of adjunct questions identical to the lesson and final test questions.

4. Instructions on to-be-learned material can be as effective as practice on to-be-learned material alone; however, both together are the most effective.

Note. This work was accomplished under FY80-82 work unit ZR000-01-042-06-01.01: Enhancement of Information Acquisition and Storage. A more extensive discussion of the work unit is provided in Appendix B.

Delayed Feedback in Acquisition and Retention

Problem and Objective

The Navy's computer-managed instructional system, the personalized system of instruction (PSI), and precision teaching are among several instructional systems that provide immediate feedback to maximize student learning. However, immediate feedback can be expensive either in student time, instructor or proctor time, or both, and evidence is mounting that delayed feedback produces equal learning and frequently superior retention, at least when multiple-choice or fill-in test items are used. The objective of this series of experiments was to examine the relationship between the timing of feedback and long-term knowledge retention under classroom conditions.

Approach

Three experiments were conducted in courses taught according to PSI principles. Experiment I examined retention as a function of feedback delay interval, using short-answer essay tests. Experiment II varied feedback delay interval, informational quality of feedback, and test item type. Experiment III examined delay and item type.

Findings

Delay of feedback did not impair learning in any of the three experiments on immediate acquisition, retention, or study time, regardless of quality of feedback or test item type used. Further, subjects in the immediate feedback conditions reported no more initial errors than did delay subjects.

Conclusions and Recommendations

This series of experiments showed no evidence of superiority for either immediate or delayed feedback. Further, different types of feedback (varying the amount of information) did not produce differential levels of retention. The repeatable quizzing aspect of PSI probably makes feedback a less potent variable than it is in other types of courses, since students have to learn smaller quantities of material for each test and have many opportunities to learn from whatever type of feedback is provided. Whatever accounts for these findings, however, the expense and trouble involved in providing immediate feedback do not appear to be warranted.

Note. This work was accomplished under FY79-81 work unit ZR000-01-042-06-01.02: Delayed Feedback in Acquisition and Retention.

Improving Language-learning and Problem-solving Abilities

Background

Until recently, the assignment of recruits to schools or remedial education programs has depended almost exclusively on the use of static paper-and-pencil test items to which the examinee must respond. Although such tests are useful for general classification purposes, they provide little information that is useful in subsequent evaluation or instruction. What has been lacking is an approach that allows training to be geared toward the unique strengths and weaknesses of a given individual.

Approach

The present effort is an attempt to investigate the information-processing characteristics of individuals engaged in intellectually demanding, problem-solving tasks. Efficient problem solving is seen as the result of the programmatic use of a hierarchy of cognitive processes. The aim of this research effort was to investigate the efficiency with which different individuals use these components, thereby highlighting unique strengths and weaknesses. Once the pattern of problem solving has been ascertained, it can be used in individual selection or assignments; alternatively, an individualized program of instruction can be developed to improve learning.

Eye scan patterns were used to tap problem solving strategies. Eye fixations were tracked by means of an oculometer. Measures were taken while the subjects attempted to solve a series of analogy items taken from the Ravens Advanced Progressive Matrices Test.

Findings

Eye scan patterns can distinguish between average and low ability personnel (as measured by the ASVAB), index the degree to which individuals prefer inductive versus deductive reasoning, determine the amount of information an individual needs before hypotheses can be formulated, measure how much information can be maintained in memory while performing associated tasks, and, in general, provide much information about the mechanics of an individual's thinking.

Note. This work was accomplished under FY80-82 work unit ZR000-01-042-06-01.03: Improving Language-learning and Problem-solving Abilities. A more extensive discussion of the work unit is provided in Appendix C.

INDEPENDENT EXPLORATORY DEVELOPMENT (IED) PROGRAMS

The Independent Exploratory Development (IED) program at NAVPERSRANDCEN was initiated in FY76. In FYs 80 to 82, programs have been funded in the six areas indicated in Table 2.

Table 2

IED Funding at NAVPERSRANDCEN

	Area	Funding in Thousands of Dollars[a] FY80	FY81	FY82
ZF66-512-001				
-010	Evaluation of Psychobiological Methods	35	15	5
-030	Measurement Support for Human Factors Research	20	0	0
-050	Factors Affecting the Acceptance of Change	30	45	50
-060	Analytical Techniques for Personnel Research	40	0	0
-070	Methods for Improving Maintenance Procedures	0	15	0
-080	Production and Use of Graphics	0	50	80
Total		125	125	135

[a]As of 1 April 1982.

The IED work that has been accomplished in FY81 is described in the following pages. In addition, Appendices D and E provide extensive articles on efforts receiving substantial IED support over 2 or more years.[1] These articles, which describe research on event-related brain potentials and office technology respectively, were prepared for this report and for possible publication in appropriate periodicals.

[1]Some of the work described in Appendix D was performed with IR funding as well as IED funding.

EVALUATION OF PSYCHOBIOLOGICAL METHODS

Personnel Applications of Event-related Brain Potentials

Background

Event-related potentials (ERPs) are small electrical signals recorded from electrodes placed in contact with the scalp that represent the brain's response to systematic auditory or visual input. It is believed that most traditional paper-and-pencil tests emphasize the skills that are largely the function of the left hemisphere of the brain (in most right-handed people), while effective job performance may emphasize right hemisphere skills. Event-related potential research technology permits the recording and analysis of signals from each hemisphere separately.

Navy use of color-coded visual display systems is expected to increase in the coming years. Color coding increases the amount and complexity of information that may be displayed. Such increased sophistication in visual display systems places great demands on the personnel required to operate them. Better understanding of brain responses to color-coded displays may result in improved design of display systems and improved operator performance.

Findings

Testing of subjects in NAVPERSRANDCEN's Biotechnology Laboratory has demonstrated the effects of color and color interaction on brain function and perception. Specific visual ERP components have been found for the three basic color processes (presumably red, green, and blue). Several interesting findings emerged:

1. Visual ERP data obtained from red, green, and blue stimuli tended to maximize individual differences to a greater degree than did white stimuli.

2. Presentations of stimuli to each eye separately accentuated individual differences to a greater degree than did presentation to both eyes simultaneously.

3. Greater ERP amplitude variability was found at the back of the head (visual perception and association area) than at the front of the head where other brain functions are thought to occur.

4. Greater variability was found for red than for green or blue.

5. Greater variability was found over the left hemisphere than the right hemisphere.

6. The visual ERP data suggested that the subjects may be clustered into several groups based on analyses of the response patterns to the color stimuli.

In several projects completed prior to FY 1981 under other work units, it was shown that event-related potentials were promising predictors of performance in tasks as varied as learning to read, operating sonar equipment, or piloting an aircraft.

Note. This work was accomplished under FY81-82 work unit ZF66-512-001-010-03.02: Color-specific Visual Event-related Brain Potentials as related to Navy Display Systems. A more extensive discussion of this work unit and related IR/IED efforts is provided in Appendix D.

FACTORS AFFECTING THE ACCEPTANCE OF CHANGE

Implementation and Evaluation of Office Technology

Background

The widespread availability of computers that are increasingly cost effective has had an enormous and increasing impact on the rate of implementation of office technology. Although the potential benefits to white-collar productivity are great, there are, as with any innovation that requires significant changes in the way work is accomplished, obstacles or impediments that threaten full implementation. This is not surprising, since the introduction of computers will bring about radical changes in many office environments. To the extent that these impediments are not dealt with, the full benefits of computer technology will not be realized.

Approach

Professionals, managers, and support personnel from a Navy R&D laboratory were interviewed to identify their concerns about a new office technology system. Personnel from the Naval Military Personnel Command were also interviewed regarding a similar technological implementation.

Findings

A diversity of views was found concerning the proper role of the new technology in the office setting. This can be attributed in part to the differences in familiarity with automated office equipment and in the duties and responsibilities confronting each respondent. Despite these differences, the majority of support, managerial, and professional personnel were positively disposed toward the new system. Overall, the proposed office automation was welcomed as an opportunity to do something new and different, and its advantages outstripped any disadvantage.

Conclusions

The following are suggested to help overcome impediments or obstacles to full implementation of new technologies.

1. An effective support system for the new technology, including such specifics as the provision of funds and training materials, rearrangement of prevailing organizational structure to ensure compatibility with the function and goals of the innovation, and availability of local resource personnel to provide practical advice.

2. A well publicized organizational philosophy concerning the technology that defines the scope, function, and impact of the change on the organization and its people.

3. Socialization mechanisms geared toward closing the gap between the employee's perceptions of innovations and the views articulated by the organization.

Note. This work was accomplished under FY 81-82 work unit ZF66-512-001-050-03.08: Impediments to the Implementation of Innovation. A more extensive discussion of the work unit is provided in Appendix E.

METHODS FOR IMPROVING MAINTENANCE PROCEDURES

Remote Expertise for Maintenance

Background

Shore-based technical assistance is generally requested when a ship has determined that its own unit personnel are unable to diagnose and repair an equipment casualty. Although such casualties may constitute only a limited portion of the maintenance problems encountered in the fleet, they are especially detrimental to overall operational readiness because of the lengthy delays usually encountered before outside technical assistance is provided. One possible way to reduce the amount of time that a ship spends in a degraded state of equipment readiness at sea is to provide a reliable and effective two-way communication link between expert technicians on shore and deployed units of the fleet using modern communication satellite technology.

The technological feasibility of establishing a remote communications assistance network was recently demonstrated in a project sponsored by the Naval Sea Systems Command (NAVSEA 62C), titled "Shipboard Engineering Assistance for System Test and Repair via Satellite" (SEASTARS).[1] The SEASTARS project team was successful in demonstrating the technological compatibility of various communications accessory equipments that would support such a network, as well as the methodology to be used in integrating these units with a commercially-available communications satellite system. However, the project team did not demonstrate (either directly or through simulation) the problem-solving capabilities of this performance-aiding concept. The current study was undertaken to provide such a demonstration and to identify factors that might affect the use of such a procedure when troubleshooting Navy equipment casualties.

Approach

Equipment casualties, representative of those requiring outside technical assistance, were inserted in the AN/SQR-17 sonar detecting-ranging set in two demonstration sessions. For each session, an SQR-17 maintenance instructor from the Fleet Anti-submarine Warfare Training Center, Pacific (FLEASWTRACENPAC), San Diego served as the remote maintenance technician (or expert), while 2nd and 3rd class surface sonar technicians (STGs) from the fleet served as shipboard (or attendant) technicians. A two-way voice communication link was established between the expert and attendant technicians via an audio intercom. For session 1, a video link was also provided that allowed the expert to view those portions of the equipment that would provide information of potential relevance to the troubleshooting situation. For session 2, the expert was instructed to use the video link only in those cases where he could no longer proceed with troubleshooting based on audio information alone.

The attendant technicians were stationed at the equipment site and were responsible for reporting the symptoms associated with each casualty to the expert technician. They were also responsible for performing various fault isolation tests (using standard Navy test equipment) and for removing and replacing printed circuit cards as directed by the expert.

The expert technician was stationed in an adjacent room. He was instructed to direct the attendants verbally in troubleshooting and repairing equipment casualties and, as

---

[1] American Management Systems Incorporated. SEASTARS, Technological feasibility study and preliminary test report (Contract N00024-79-C-5683). Arlington, VA: Author, December 1980.

applicable, to direct the television camera to those portions of the sonar system or supporting test equipment.

Three troubleshooting problems were inserted in the SQR-17 for each session. The faults were identical in both cases but their order of presentation varied between sessions. For each problem, conversations between the expert and attendant technicians, as well as the total time spent in troubleshooting, were recorded.

Findings

The remote maintenance technicians were successful in solving five of the six problems inserted in the SQR-17 sonar. The problem that was not solved was terminated due to time limitations imposed by the testing schedule. Conversation time between the expert and attendant technicians accounted for an average of only 17 percent of the total time spent in troubleshooting each problem. When visual information was provided, the expert technicians were interested in viewing only limited portions of the sonar set and supporting test equipment. During session 2, an imprecise verbal description of the equipment casualty symptoms associated with one troubleshooting problem resulted in the expert technician pursuing a lengthy and ineffective troubleshooting strategy. When the expert requested video information to confirm the casualty symptoms, he was able to locate and solve the problem in less than 30 minutes.

Conclusions

To the extent that the current study was able to simulate the conditions under which remote communications assistance would be provided, the feasibility of this concept was successfully demonstrated. The large proportion of time spent in nonconversational activity for each troubleshooting problem indicates that the communications link could be used intermittantly during the course of a remote communications assistance event. Expert maintenance technicians must be provided with a precise verbal description of all relevant symptoms of an equipment casualty to facilitate timely and effective troubleshooting strategies when only audio link remote communications assistance is employed.

Note. This work was accomplished under FY81 work unit ZF66-512-001-070-03.07: Remote Expertise for Maintenance. A more extensive discussion is presented in NPRDC SR 82-15 (Nugent) (AD-A111 525).

PRODUCTION AND USE OF GRAPHICS

Theory of Graphic Representation

Background

The adage that a picture is worth a thousand words is not invariably true. Although graphics are often useful, they can be confusing and thus interfere with learning and understanding. How can we determine when and how graphics will be effective in instruction?

Approach

Two Navy tasks were identified that presented unusual difficulty for students and that seemed amenable to solution by creating effective graphic instructional materials. In the first study, a concrete analogy was developed (using poker chips) that proved easy for students to understand and that effectively represented the relationships in a simple electronic circuit. The poker chip model is more coherent and integrated than the abstract, theoretical, charged-particle representation currently used. The effectiveness of this representation is being tested by comparing the progress of groups taught how to solve Ohms law problems using either the actual course materials or using the newly developed more concrete representation. Two groups of students received parallel forms of instruction using booklets. Evaluation of training material effectiveness includes time through the booklets, performance on a set of problems, and performance in course lessons. Data from this experiment are being analyzed. Based on these results, a program will be developed on the Terak microcomputer that will simulate the circuit problems found in the Basic Electronics/Electricity (BE/E) course. The feasibility of using these approaches for training and assessing student performance will be tested in FY82.

The second effort involved developing microcomputer graphic techniques for training qualitative understanding of relative motion in maneuvering. A preliminary study was undertaken to begin evaluation of this training. The interactive computer aided instruction (CAI) system provides students with a conceptual model of the relationships between real-world events and elements of the maneuvering board plot. The system developed is a self-contained software package that runs on Terak microcomputer systems installed at the Fleet Combat Training Center Pacific. During the last part of FY81, the system was evaluated at the OS "C" school. In FY82, plans are to develop an intelligent automated tutorial facility for the maneuvering board domain.

Findings

In both studies, preliminary results suggest that the graphic representations had the desired effects. The concrete analogy is useful for teaching difficult conceptual materials in basic electronics. Analysis of problems confronted by the researchers in using the maneuvering board and teaching students to use the board led to the development of computer programs that present a unique graphic representation to students. This presentation is intended to help students conceptualize the relation between relative motion and geographic plots of the same information. Preliminary evaluation during the latter part of FY81 suggests that student learning can be aided with this system. Instructor acceptance is enthusiastic.

Note. This work was accomplished under FY81-82 work unit ZF66-512-001-080-03.08: Theory of Graphic Representation.

Judging the Comprehensibility of Technical Graphics

Problem

Technical manuals used by Navy personnel in operating and maintaining complex equipment may be composed of 50-60 percent or more graphics. Guidelines for ensuring the comprehensibility of such graphics are virtually nonexistent, except for very intuitive "helpful hints" to illustrators and technical writers. One reason for this lack is the absence of information about the dimensions of graphics that contribute to their usefulness. A first step in the identification of such dimensions is to have experts (in this case, senior personnel who have reason to use a specific kind of graphic) rate graphics according to those dimensions of importance to them.

Approach

The objective of the present research was to examine the ability of experienced Navy electronic technicians to differentiate reliably a "good" graphic from a "bad" graphic, where "good" and "bad" refer to the usability of a graphic in a simulated job situation. The specific type of graphic selected for this first step in the investigation was the electronic schematic. A set of representative schematics was assembled that intuitively appeared to incorporate all the relevant dimensions that might have an effect on their usability. The resulting set of schematics was presented to 40 senior Navy electronic technicians on an individual basis. Each technician was asked to sort the schematics into six categories, from one (very good) to six (very bad), according to his perception of each schematic's usability in a trouble-shooting situation that was described to him by the experimenter. Two such trials were conducted by each subject. The subject was not required to provide a rationale for his ratings. Comments made by the subject, however, were recorded.

The data were analyzed to determine inter- and intrarater agreement, mean judgments of schematic quality, and groupings of schematics and experts.

Findings

Perhaps the major conclusion to be drawn from this investigation is that experts (in this case, senior electronic technicians) are indeed able to differentiate reliably high from low quality in schematics. We may be much more confident of the validity of expert ratings in this type of task in developing training materials and in future research.

It is tentatively concluded that the expert raters are able to respond differentially to difficult aspects or dimensions of schematics, and that the methods developed will permit the elucidation of these dimensions and the confirmation of the dimensions derived from the cluster analysis of the present data.

Unfortunately, the expert raters proved to be somewhat taciturn and did not provide as much commentary information as had been anticipated, thus limiting the usefulness of verbal data collected. More effort will be expended in future work toward soliciting comments from the experts.

Note. This work was accomplished under FY81-82 work unit ZF66-512-001-080-03.09: Producing Effective Technical Graphics.

IR AND IED WORK UNITS FOR FY81-82

Independent Research			Dollars (in thousands) FY81	FY82[a]
ZR000-01-042-04:	SELECTION AND RETENTION			
-01	Measurement Theory and Methods in Personnel Research		83	68
-01.03	Differential Worth for Optimal Individual Placement	Dr. L. Kroeker Code 12 A/V: 933-2176	10	5
-01.04	Unobtrusive Measures of Attitude	Dr. C. A. Robinson Code 14 A/V: 933-7122	12	0
-01.05	Methods for Clustering Tasks	Mr. R. Chatfield Code 12 A/V: 933-2176	15	0
-01.06	Models for Personnel Management	Mr. J. Silverman Code 11 A/V: 933-6384	40	30
-01.07	Models for Calibrating Multiple-choice Items	Dr. J. Sympson Code 12 A/V: 933-2176	6	0
-01.08	Cognitive Speed Tests	Dr. B. Rimland Code 01C A/V: 933-6122	0	21
-01.09	Adaptive Ability Tests	Dr. D. Wetzel Code 12 A/V: 933-2176	0	3
-01.10	Models and Measures of Human Performance	Dr. R. Sorenson Code 03 A/V: 933-2231	0	9
-03	Retention of Quality Personnel		95	0
-03.01	Methods for Projecting Petty Officer Retention	Dr. E. Curtis Code 16 A/V: 933-2396	65	0
-03.02	The Reenlistment Decision Process	Dr. L. Doherty Code 301 A/V: 933-2231	30	0

[a]Funding as of 1 April 1982.

Independent Research (Continued)			Dollars (in thousands) FY81	FY82[a]
-042-06:	LEARNING AND TRAINING			
-01	Instructional Psychology		77	127
-01.01	Enhancement of Information Acquisition and Storage	Dr. J. Ellis Code 13 A/V: 933-7121	20	42
-01.02	Delayed Feedback in Acquisition and Retention	Dr. N. Van Matre Code 14 A/V: 933-7122	13	0
-01.03	Improving Language-learning and Problem-solving Abilities	Dr. R. Wisher Code 13 A/V: 933-6400	44	5
-01.04	Metaphores and Analogies in Training	Dr. S. Hearold Code 15 A/V: 933-2371	0	30
-01.05	Cognitive Storage Mechanisms and Learning to Read	Dr. M. Baker Code 15 A/V: 933-6803	0	50
-042-08:	GROUP BEHAVIOR			
-03	Group Performance Effectiveness		0	65
-03.01	Group Problem Solving in Operational Settings	Dr. W. Montague Dr. E. Hutchins, Jr. Code 13 A/V: 933-7121	0	15
-03.02	Organizational Effectiveness of Program Management Offices	Mr. T. Enderwick Mr. T. Koslowski Code 16 A/V: 933-6935	0	50
TOTAL			255	260

[a]Funding as of 1 April 1982.

Independent Exploratory Development

			Dollars (in thousands) FY81	FY82[a]
ZF66-512-001:	IED SUPPORT TECHNOLOGY			
-010	Evaluation of Psychobiological Methods		15	5
-010-03.02	Color-specific Visual Event-related Brain Potentials as Related To Navy Display Systems	Dr. G. Lewis Code 17 A/V: 933-6617	15	5
-050	Factors Affecting the Acceptance of Change		45	50
-050-03.08	Impediments to the Implementation of Innovation	Dr. J. Sheposh Ms. V. Hulton Code 16 A/V: 933-6935	45	50
-070	Methods for Improving Maintenance Procedures		15	0
-070-03.07	Remote Expertise for Maintenance	Mr. W. Nugent Code 17 A/V: 933-2371	15	0
-080	Production and Use of Graphics		50	80
-080-03.08	Theory of Graphic Representation	Dr. W. Montague Dr. J. Hollan Code 13 A/V: 933-7121	35	40
-080-03.09	Producing Effective Technical Graphics	Dr. T. Curran Code 13 A/V: 933-6400	15	40
TOTAL			125	135

[a]Funding as of 1 April 1982.

IR/IED PUBLICATIONS

Dillon, R. F., & Wisher, R. A. The predictive validity of eye movement indices for technical school quality test performance. Applied Psychological Measurement, 1981, 5, 43-49.

Ellis, J. A., Wulfeck, W., II, & Montague, W. E. The effect of adjunct and test questions similarity on study behavior and learning in a training course. American Educational Research Journal, 1980, 17, 449-457.

Federico, P-A. Domain of processing and recognition memory for shapes. Bulletin of the Psychonomic Society, 1980, 16, 261-264.

Federico, P-A. Some effects of encoding, codability, and exposure upon recognition memory. Bulletin of the Psychonomic Society, 1980, 16, 89-92.

Hutchins, E., & McCandless, T. P. MANBOARD: A graphic display program for training relative motion concepts (NPRDC Tech. Note 82-10). San Diego: Navy Personnel Research and Development Center, March 1982.

Laabs, G. W. Perceptual processing in motor memory. In Nadeau, C. H., Halliwell, W. R., Newell, K. M., & Roberts, G. C. (Eds.). Psychology of Motor Behavior and Sport--1979. Champaign, IL: Human Kinetics, 1980.

Laabs, G. W., & Simmons, R. W. Motor memory. In D. H. Holding (Ed.). Human Skills. London: Wiley, 1981.

Lau, A. W., Newman, A., & Broedling, L. A. The nature of managerial work in the public sector. Public Administration Review, 1980, 40, 513-520.

Lau, A. W., & Pavett, C. M. The nature of managerial work: A comparison public and private sector managers. Group and Organization of Studies, 1980, 5, 453-466.

Lau, A. W., Pavett, C. M., & Newman, A. Public and private sector managers: Are they really that different? (NPRDC Tech. Rep. 82-41). San Diego: Navy Personnel Research and Development Center, April 1982. (AD-A114 021)

Lewis, G. W., & Froning, J. N. Sensory interaction, brain activity, and reading ability in young adults. International Journal of Neuroscience, 1981, 15, 129-140.

Lewis, G., & Rimland, B. Psychobiological measures as predictors of sonar operator performance (NPRDC Tech. Rep. 80-26). San Diego: Navy Personnel Research and Development Center, May 1980. (AD-A085 030)

Montague, W. Is simulation fidelity the question? (NPRDC Tech. Note 82-13). San Diego: Navy Personnel Research and Development Center, April 1982.

Moonan, W. J. Tools of analysis and decision. Volume IV: On measurement and ANOVA. San Diego: Navy Personnel Research and Development Center, February 1981.

Naitoh, P., & Lewis, G. W. Statistical analysis of extracted features. In Yamaguchi, N., & Fujisawa, K., Recent advances in EEG and EMG data processing. Proceedings of the International Conference on EEG and EMG Data Processing, Kanazawa, Japan, 10-12 September 1981.

Newman, D. Test of a decision model as a decision-making aid for human factors design problems (NPRDC Tech. Note 82-5). San Diego: Navy Personnel Research and Development Center, February 1982.

Nugent, W. A. Examining the feasibility of a remote communication assistance network (NPRDC Spec. Rep. 82-15). San Diego: Navy Personnel Research and Development Center, February 1982. (AD-A111 525)

Pass, J. J., & Chatfield, R. E. Empirical comparison of binary and continuous proximity measures for clustering occupational task data (NPRDC Tech. Rep. 82-36). San Diego: Navy Personnel Research and Development Center, March 1982. (AD-A112 930).

Pavett, C., & Lau, A. An examination of research and development managerial jobs across two sectors. Journal of Management Studies (in review).

Pavett, C., & Lau, A. The influence of hierarchical level and functional specialty. Academy of Management Journal (in press).

Rimland, B. Independent research and independent exploratory development at the Navy Personnel Research and Development Center--FY80 (NPRDC Spec. Rep. 81-20). San Diego: Navy Personnel Research and Development Center, June 1981.

Robinson, C. A., Tomblin, E. A., & Houston, A. Computer-managed instruction in Navy technical training: An attitude survey (NPRDC Tech. Rep. 82-19). San Diego: Navy Personnel Research and Development Center, December 1981. (AD-A109 664)

Snow, R. W., Federico, P-A., & Montague, W. E. (Eds.). Conference Proceedings: Aptitude, learning, and instruction (Vols. 1 and 2) (NPRDC/ONR Tech. Rep. 81-5). San Diego: Navy Personnel Research and Development Center, January 1981. (AD-A099 209 and AD-A099 208)

Sorenson, R. C. Independent research and independent exploratory development at the Navy Personnel Research and Development Center (Spec. Rep. 80-23). San Diego: Navy Personnel Research and Development Center, June 1980).

Sturges, P. T., Ellis, J. A., & Wulfeck, W. H. Effects of performance-oriented text upon long-term retention of factual material (NPRDC Tech. Rep. 81-22). San Diego: Navy Personnel Research and Development Center, September 1981. (AD-A106 096)

IR/IED PRESENTATIONS

Ellis, J., & Wulfeck, W. Quality control of instruction and instructional development. American Psychological Association, Los Angeles: August 1981.

Ellis, J., Wulfeck, W., Montague, W., & Fredericks, P. The effects of instructions and practice on learning. American Educational Research Association, Los Angeles: April 1981.

Federico, P-A., & Landis, D. B. Are cognitive styles independent of abilities and aptitudes? Meeting of the American Psychological Association, Montreal: August, 1980.

Federico, P-A. Individual differences and mastery learning in computer-managed instruction. American Educational Research Association, Los Angeles: April 1981.

Lau, A. The nature of managerial work: Implications for selection, development and performance appraisal. International Association for Personnel Women, San Diego: 1981

Lau, A., & Pavett, C. What's different about R&D managerial jobs? National Academy of Management, San Diego: August 1981.

Lewis, G. W. Event-related brain potentials and job performance. 24th Annual Human Factors Society Meeting, Los Angeles: 13-17 October 1980.

Montague, W. E. A common flaw in research design: Inconsistency between learning and testing requirements. Annual Meeting of the American Educational Research Association, Boston: April 1980.

Montague, W. E., Is simulation fidelity the question? Army Research Institute, Washington, DC: July 1981.

Montague, W., Ellis, J., & Wulfeck, W. After years of instructional research, do we know more than grandma did about how to teach people? American Educational Research Association, Los Angeles: April 1981.

Pass, J., & Chatfield, R. Methods for clustering occupational task analysis data. Air Force Human Resources Laboratory, San Antonio, TX: April 1981.

Pavett, C., & Lau, A. Mintzberg revisited: A new look at the influence of contingency variables on the nature of managerial work. Western Academy of Management, Monterey, CA: April 1981.

Sturges, P., Ellis, J., & Wulfeck, W. Effects of performance-oriented text on longterm retention of factual material. American Educational Research Association, Los Angeles: April 1981.

Wetzel, S., Ellis, J., Wulfeck, W., & Konoske, P. Controlling test comprehension: Adjunct questions vs. instruction. American Educational Research Association, Los Angeles: April 1981.

APPENDIX A

A PROCEDURE TO REVISE ESTIMATES OF PSYCHOLOGICAL SCALE VALUES

A PROCEDURE TO REVISE ESTIMATES OF PSYCHOLOGICAL SCALE VALUES

Leonard Kroeker

Personnel assignment policies affect almost every facet of naval operations. Therefore, decisions to alter or discontinue particular policies must be made with due consideration given to the potential effects of such changes, the conditions under which the effects may occur, and the state of prevailing operational conditions. One aspect of this requirement involves the capability to adjust process parameter estimates so that new information concerning present and future conditions may be optimally integrated. This capability is of particular interest to those dealing with dynamic operational systems such as the Navy's optimal-sequential assignment procedure, CLASP (Classification and Assignment within PRIDE (for Personalized Recruiting for Immediate and Delayed Enlistment)).

The purpose of this article is to describe a tool that may be used by policy makers to alter estimates characterizing a process, based on more recently acquired data elements. The original parameter estimates are typically derived from analyses of human judgments and revision is often motivated by the fact that recent data more accurately reflect system status and frequently represent data whose quality is superior to the original.

This article addresses two questions: (1) Why would a decision maker elect to use such an aid and (2) under what conditions would it be most beneficial? An example drawn from an operational setting may be helpful in answering the first question. Consider the case of a manpower manager who has assessed billet types with respect to anticipated losses at the end of the first duty tour. Typically, the data exist in the form of an ordering among objects with respect to a loosely defined criterion; in this example, billet types are ordered. The ordering is often the result of applying a primitive measure such as a ranking procedure. Regardless of the rule applied, decisions are needed concerning which position estimates are to be modified and to what extent they should be adjusted. The latter requires a systematic approach to stimulus location estimation. The automated revision capability that was developed enables managers to (1) replace a subjective approach with an objective one, (2) derive refined estimates by incorporating partial data sets, and (3) assess the adequacy of the revision and plan for future modification.

The iterative character of the solution process permits partial data sets, usually generated by more recent sets of observations than the original, to be incorporated in stages. At each stage, the solution is influenced by the new data and the latter determines which parameter estimates will be affected and to what extent. When the true parameter values are stable but unknown, the incoming information serves to improve solution quality. Thus, the specific purposes of the research project were:

1. To formulate a mathematical model expressly designed to integrate partial data sets into a paired-comparison solution to derive better stimulus location estimates.

2. To develop specifications for the use of the model in refining estimates and planning future modifications.

3. To conduct a preliminary evaluation study of model operation under various conditions.

Problem Context

Policy makers continually face problems that involve the setting of priorities and the revision of those priorities when circumstances change. The processes underlying the revision are not well understood and algorithms, if they exist, are rarely employed. Descriptions of the processes followed in making adjustments are difficult to obtain since the decision processes that can be investigated in the laboratory involve numerous artificial constraints and field investigations are fraught with methodological difficulties. As a result, revision is based on intuitive grounds with no formal mechanism employed to determine which information ought to be included, how salient the information sources may be, or how the information sources ought to be weighed to achieve an optimal mixture.

The question of interest in this paper involves the translation of the intuitive process to that of a systematic procedure that can be used routinely for priority revision problems. In order to use the procedure, it must be placed in the appropriate form. The manner in which this is done is discussed in the following sections of the paper.

The problem was first posed when attempting to modify scale values describing the relative positions of job categories in a complexity scale. It grew out of a study involving the assignment of first-term enlistees to entry level Navy jobs. In particular, an assignment algorithm capable of handling personnel one at a time was being developed. One of the policy functions included in the algorithm required an up-to-date ordering of jobs with respect to complexity level.

The utility model employed to match recruit applicants with entry level Navy jobs is called CLASP. This model is represented in Figure A-1 as a set of functions that produce an index that reflects the quality of any given assignment. It consists of five utility generating functions that quantify different aspects of the assignment decision, allowing information concerning applicant background data, aptitude test scores, personal job preferences, and Navy requirements to be integrated in a systematic, meaningful way. Several utility model components involve empirically-derived scales that require modification in response to changing conditions in the military acquisitioning process. The complexity scale found in component D is an example of a scale requiring such modifications. Navy jobs are ordered (along the complexity continuum) and changes in the ordering affect the utility values that are subsequently derived.

The empirically-derived scales consist of stimuli scaled using a paired comparison procedure. Frequently, data gathered in this way are imprecise or incomplete. The former is readily apparent when one considers the variability in human judgments and the latter when one considers the large number of comparisons demanded by a moderately-sized stimulus set. Revision of the initial scale values is almost always necessary and it must often be accomplished in stages as new data become available.

Background Research Literature

The scaling method to be used in this paper is that of paired comparisons, a process for assigning unidimensional scale values to a set of objects. A unidimensional scale can be obtained by applying Thurstone's law of comparative judgment to frequency data. The procedure is well described elsewhere (Torgerson, 1958; Bock & Jones, 1968; Arbuckle & Nugent, 1973). The standard scaling procedure involves the use of data obtained from all pairs of objects in the set; in other words, a complete data matrix is analyzed.

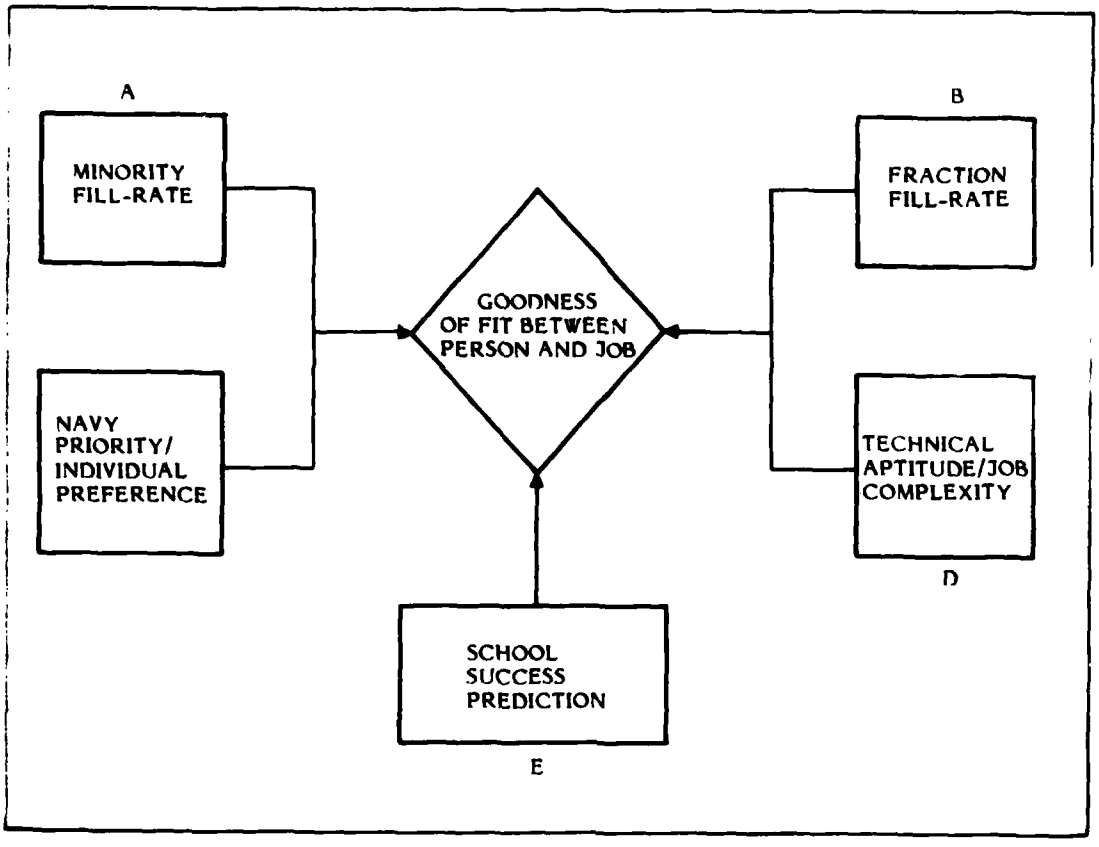


Figure A-1. Personnel assignment utility model.

Gulliksen (1956) has derived an incomplete data solution for paired comparison data although it is not entirely clear under which conditions his method will yield a unique solution. Schonemann (1970) recommends a slight change in procedure that renders the scale values obtained with Gulliksen's solution unique up to an additive constant. As a result, Gulliksen's solution is remarkably tolerant in its acceptance of empty cells in the paired comparison matrix.

More recently, scaling procedures governing the assignment of psychometric scale values from incomplete paired comparison matrices have been investigated by Indow and Ida (1975). They propose a method that is almost equivalent to one developed by Morrissey (1955). The motivation to investigate the incomplete case stems in part from the fact that it is practically impossible to obtain a complete paired comparison matrix when the number of stimuli is large.

Since the research in this study involves the use of partial data sets, the research literature dealing with both complete and incomplete matrices is relevant. The solution strategy to be proposed will call for the establishment of an original paired comparison matrix regardless of the quality of its constituent parts. The quality of the matrix and, hence, the solution will be subsequently upgraded in an iterative fashion. No literature exists on this approach to the incomplete data question.

The Model

This section describes the operation of the revision process. For this purpose, it is useful to consider certain structural aspects of the underlying model. Although the structure is best communicated by a mathematical expression, it is described here in a nonquantitative format.

Since the modification process involves the alteration of a set of scale values, consider a set of n stimuli that are described by numerical values. The latter represent a measure of the amount of a specified attribute possessed by the stimuli.

Usually the values corresponding to the stimuli are unknown and data must be gathered and analyzed to obtain them. The stimuli are compared pairwise by persons who are asked to determine which stimulus of the pair exhibits the greater amount of the attribute and to what extent this dominance is manifested.

To a significant degree, the nature of the judgmental data determines the characteristics of the resulting scale. The latter will be useful in ordering the stimuli but not necessarily useful in reflecting absolute measure. In other words, statements concerning absolute amounts of the attribute will have no meaning.

A psychological discrimination process underlies the paired comparison procedure. It is assumed that the individual making the judgment is responding as though each stimulus lies on a continuum that represents the attribute in question. The objective of the scaling exercise is to discover the relative positional value corresponding to each stimulus. The determination of this value is difficult because the measurement process is contaminated by a stochastic noise component. The latter is the term used to describe the result of statistical fluctuations superimposed on consistent judgments. Thus, on most occasions, the measurement process results in a value that differs from the true positional value of the stimulus. If an indefinitely large number of measurements is taken, the result is a distribution of values around the true position of the stimulus.

The attempt to retrieve information about the relative positions of a pair of stimuli is complicated by the statistical character of the stimulus position estimate. For example, when two stimuli lie close to one another on the continuum, it is possible for an erroneous judgment to be made concerning which of the two exhibits the greater amount of the attribute.

The potential errors can be handled rather easily in the solution procedure. Knowledge of the relative position of any given stimulus is obtained through each of the possible pairwise judgments involving the stimulus. If a pair of stimuli lie next to one another on the continuum, approximately half the judgments will reflect the correct order and half the incorrect one. As the distance between the stimuli increases, fewer incorrect judgments appear in the data. In a sense, a cancellation affect is operating to ensure that, on the average, a reliable measure of the distance between stimuli is obtained.

The judgment data are transformed and combined to form a scale with useful properties. The scale consists of stimulus location estimates corresponding to the stimuli in the set. The estimates are relatively robust since they have been derived from such a large number of data elements; namely, all possible pairwise judgments involving the stimuli.

The paired comparison matrix within which the judgmental data elements are located is identified in Figure A-2 and is used to derive the stimulus location estimates mentioned

above. To determine the need for the process described in this paper, it is necessary to examine the estimates and decide whether or not they require revision. If so, the transformed matrix in Figure A-2 is subjected to several mathematical operations that result in a smoothed data matrix. In other words, the interrelationships among the location estimates are folded back into the data matrix so that all the paired comparisons are constrained to reflect the interrelationships.

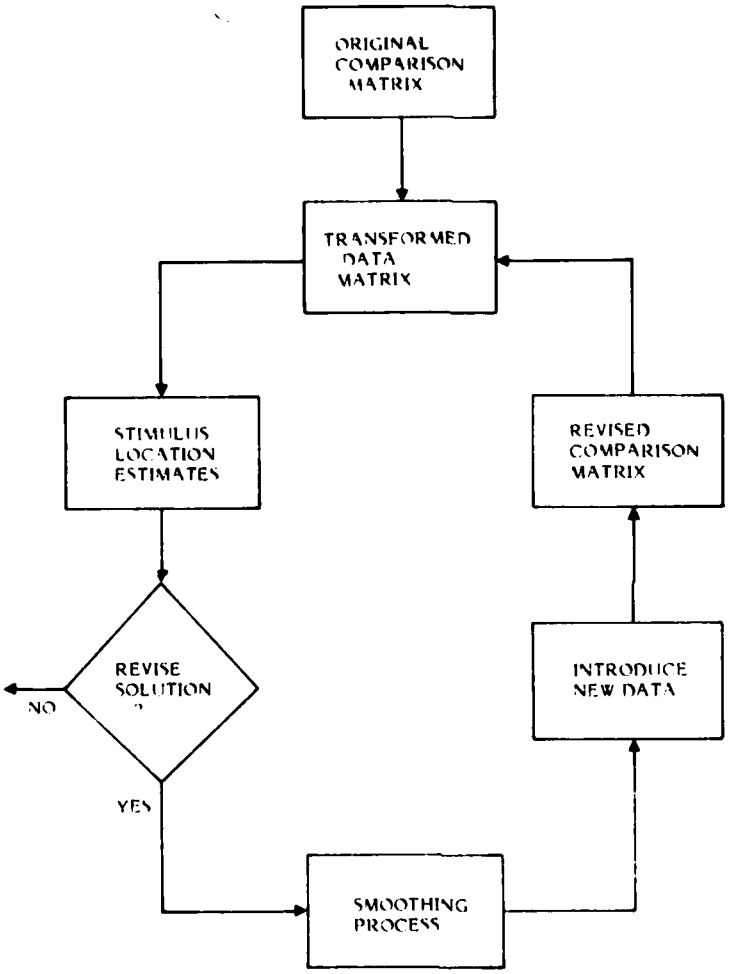


Figure A-2. Scale value modification procedure.

A data matrix that has not been subjected to the smoothing process contains judgmental data elements that are inconsistent with one another. A common illustration involves a subset of three stimuli (say, A, B, and C) that yield an intransitivity; for example, A>B, B>C, and C>A. Clearly, the first two relationships considered jointly are logically inconsistent with the remaining one. In other words, the raw data matrix contains data elements whose level of validity differs and whose quality, therefore, varies.

The smoothing process is applied to construct a matrix whose elements are of uniform quality. All portions of the resulting matrix contain information of equal worth concerning the relative positions occupied by the stimuli.

All information concerning the relative position of a specific stimulus is found within one row of the matrix. If a portion of the data in the row is destroyed, the remaining entries may be used to reconstruct the location estimate. Because of the smoothing process, a large number of the data elements within the row can be destroyed without appreciably altering the positional value of the location estimate.

Figure A-3 illustrates the process of integrating new or more recently gathered data for the purpose of updating the location estimates. In case 1, the smoothing process yields entries in row A as shown and a stimulus location estimate, B, is calculated using the entries as input data. The value of the estimate, B, does not differ from the one calculated using the raw data.

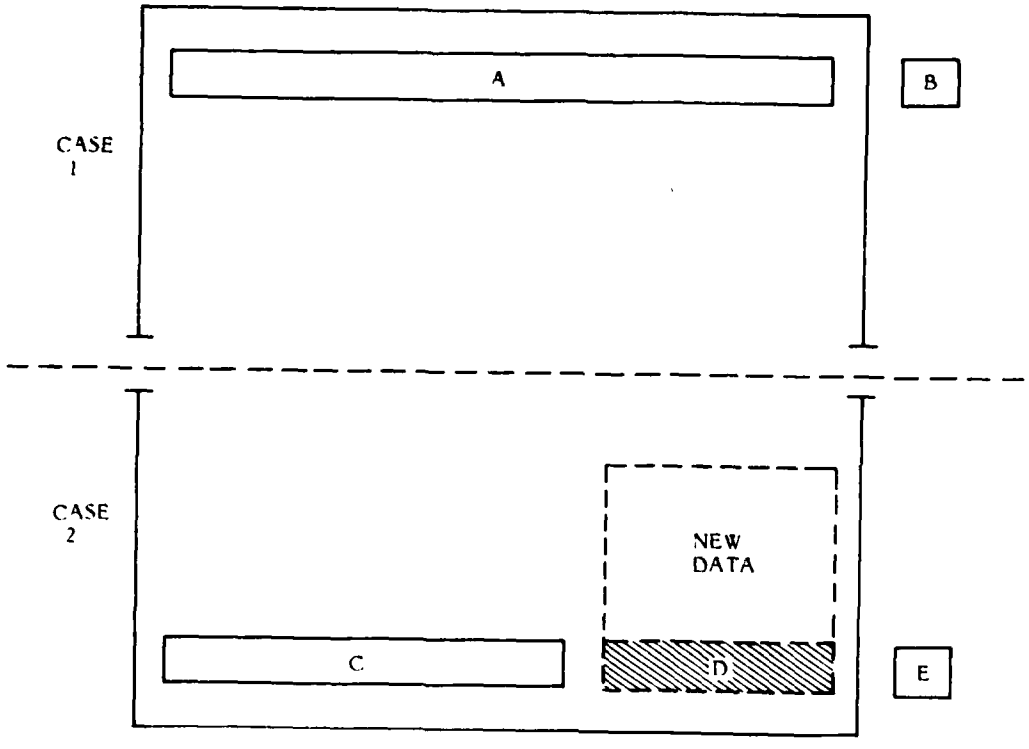


Figure A-3. Integrating new data.

However, in case 2, it is assumed that the smoothing process has been applied to the original matrix. In addition, the smoothed data elements in segment D have been replaced. The new data shown in segment D must be integrated with the smoothed data elements in segment C. The new stimulus location estimate, E, depends on the elements located in segment C and in segment D. Therefore, the estimate, E, differs from the one calculated using only the raw data.

The introduction of new data leading to a revised comparison matrix and ultimately to a new solution is shown in Figure A-2. New data sets may be introduced in stages to effect further revision of the stimulus location estimates. The degree to which the estimates are altered by the introduction of new data depends on the amount of the latter

compared to the original data. This question is examined in greater detail in a subsequent section of the paper.

Discussion of Procedure

In this section, the significance of the procedure is discussed. In routine application, the model can be expected to yield an updated version of the stimulus location estimates without having to resort to a comprehensive data collection effort. The latter is usually a massive undertaking because the number of judgments required for a problem of moderate size is so large. The advantage of using the technique is that smaller data sets, more readily obtainable by practitioners, can be easily integrated with previously collected data. In other words, one may utilize new data as they become available to improve the quality of the scale value estimates describing stimulus positions.

Occasionally, the user may discover that certain portions of the original data set are unreliable and may, therefore, employ a data acquisition plan that is limited in scope. The purpose is to cover only a limited portion of the entire data matrix to conserve resources. The data collected is then integrated and the result is a set of improved estimates.

The most obvious limitation of the procedure is its conservative approach. To appreciably alter the original estimates, large amounts of new data are necessary. In fact, if one has reason to question the validity of the original data set, it may be more economical to schedule a comprehensive new data collection effort and to derive a new solution with no reference to the original data elements because the procedure is so conservative with respect to integrating new information.

It appears that the procedure is most useful in those situations where minor modifications are needed. The extent of its usefulness under different conditions can be established through empirical studies.

The following example is an illustration of the application of the technique. During the development of the priority index used by CLASP, 20 Navy officers involved with managing enlisted ratings participated in a data collection exercise devoted to scale value revision for 86 ratings. Each officer dealt with a selected subset of ratings; namely, those with which he or she was most familiar. In each case, the list was augmented by several additional ratings chosen for experimental purposes. The judgmental data sets produced by the officers were integrated with the original data and the priority index values were modified accordingly. The resulting index is now used on a daily basis by the CLASP model in its operations involving the classification of nearly 100,000 recruit applicants per year.

It is anticipated that the procedure will usually be applied under the conditions described above. For example, suppose that Navy manpower managers wished to revise the attrition severity index (Griffin, Thomas, Euske, & Elster, 1981) developed for use in CLASP. Faced with the task of collecting judgmental data for all of the enlisted ratings, it is unlikely that one would choose to discard the existing index. Instead, one might choose to collect data involving those subsets of ratings whose positional values were considered to be in greatest need of revision and to modify the estimates accordingly.

A potentially troublesome question that must be addressed whenever one prepares to upgrade the quality of stimulus location estimates is one involving a choice between two alternative actions; namely, whether one should collect a complete data set involving all possible pairs of stimuli or several smaller data subsets. An assessment of the strengths

and weaknesses of the scale produced by the original data set is needed to determine the specific purpose of the proposed revision. For example, a scale found to be inadequate may require replacement by one derived from a complete new data set. The costs and benefits associated with either option must be estimated and the tradeoffs evaluated so that the question can be satisfactorily resolved.

Properties

This section deals with the properties of the procedure investigated thus far. Solution stability is examined first. A design to determine the effects of systematic destruction or distortion of portions of the data matrix on solution stability was developed. It was of interest to discover the degree to which the resulting scale could be altered by inserting data of poor quality.

The simulation design is represented in Figure A-4. The path on the right represents the operations employed in deriving the usual solution scale values from the original data. The path on the left represents a sequence of processes that result in the alteration of the data matrix and the subsequent modifications of the solution scale values so that they differ from the ones obtained by following the path on the right. In particular, the left path involves the substitution of random numerical values into specific portions of the matrix to which the smoothing operation has been applied.

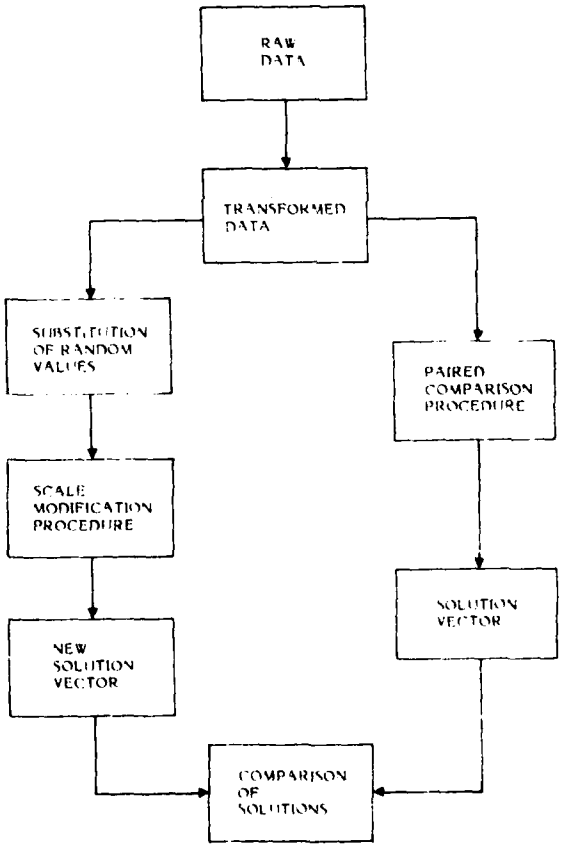


Figure A-4. Simulation design.

The result of following either path is the production of a set of scale values. The two sets are compared to determine their degree of correspondence. Deterioration in solution quality due to the insertion of the random data points is measured in terms of the degree to which the resulting solution differs from the original solution. Specifically, the correlation coefficient between the two sets of scale values is the measure used to assess the magnitude of solution deterioration.

The simulation computer program destroys a designated portion of the data matrix, inserts the new data, and compares the resulting solution with the original one. This cycle is repeated a thousand times using different random values each time but keeping all process parameters constant. The resulting correlation coefficients form a unimodel distribution that is nearly symmetric in shape. The mean of the distribution is a useful estimate of the magnitude of the effect due to the destruction of the fixed portion of the matrix.

Simulation sequences such as the one described above were performed varying the proportion of the data matrix destroyed. The amount of original data destroyed was systematically varied, ranging from a small fraction (6.7%) to nearly half the matrix. In all cases, random data were substituted for the data that were eliminated.

The effects of introducing varying amounts of random data into matrices of different sizes are shown in Figure A-5. The abscissa denotes the proportion of the matrix destroyed and the ordinate displays solution quality, as measured by the average correlation described above.

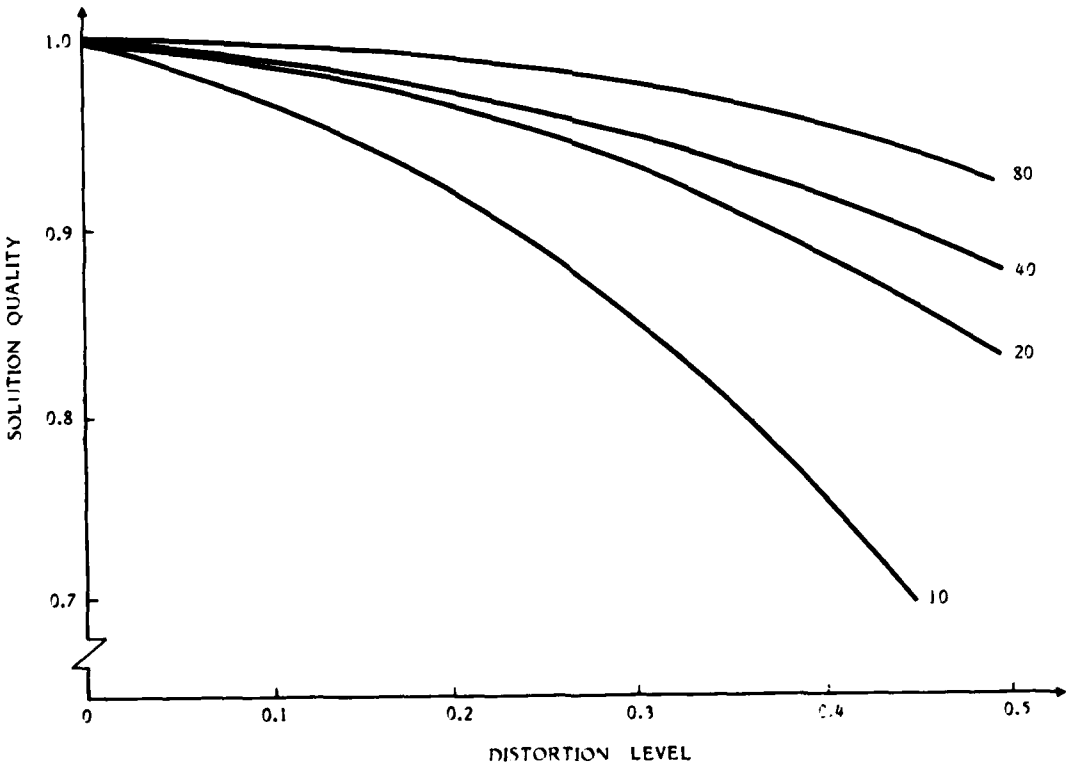


Figure A-5. Solution deterioration.

For matrices of all sizes, there is deterioration in solution quality as larger portions of the data matrix are destroyed. The effect becomes more pronounced as the distortion level is increased and as the size of the data matrix is decreased. Solution quality decreases dramatically for the 10x10 matrix when the distortion level approaches 0.5, the point at which 50 percent of the original data points have been replaced. However, it decreases only moderately for the 80x80 matrix at the same distortion level. In general, solution stability holds up well for all but the smallest matrices. It is particularly impressive when large matrices are used.

Another manipulation involved the substitution of constants as opposed to random numerical values. Under all conditions, the effect of this manipulation on the solution was negligible. In other words, the solution derived from the data set that included the substituted constants correlated with the original solution almost perfectly.

Among other characteristics, process convergence properties remain to be investigated. Imperfect data concerning relationships among stimuli generate several research questions that are noteworthy. For example, to what extent will each iteration of the revision process succeed in producing better estimates of the unknown scale values? Further, how much new data must be integrated into the original data set before convergence is achieved? Finally, a convergence criterion must be defined and a data quality measure must be derived and its properties studied.

Summary and Conclusion

This paper has described a process that can be used to revise stimulus location estimates. Its characteristics have been examined and special emphasis has been placed on investigating solution stability. Results show that solution quality deteriorates as a function of the amount of random information inserted into the basic data matrix.

In general, the process is conservative with respect to altering the positions of the location estimates. Therefore, it requires the use of large amounts of new data if substantial scale revision is desired. This is a consequence of the smoothing process that forces a systematic revision of the paired comparison relationships to conform to the relationships exhibited among the stimulus scale values obtained in the original solution.

The process is applicable to scales that require revision using only limited amounts of new data. It is assumed that an initial paired comparison solution has been derived and is available as input to the procedure.

****

Arbuckle, L., & Nugent, J. H. A general procedure for parameter estimation for the law of comparative judgment. British Journal of Mathematical and Statistical Psychology, 1973, 26, 240-260.

Bock, R. D., & Jones, L. V. The measurement of prediction and choice. San Francisco: Holden-Day, 1968.

Griffin, P., Thomas, G., Euske, K., & Elster, R. Development of an attrition severity index for CLASP (NPRDC Unpublished Manuscript). San Diego: Navy Personnel Research and Development Center, 1981.

Gulliksen, H. A least squares solution for paired comparisons with incomplete data. Psychometrika, 1956, 21, 125-134.

Indow, T., & Ida, M. On scaling from incomplete paired comparison matrix. Japanese Psychological Research, 1975, 17(2), 98-105.

Morrissey, J. H. A new method for the assignment of psychometric scale values from incomplete paired comparisons. Journal of the Optical Society of America, 1955, 45, 373-378.

Schonemann, P. H. A note on Gulliksen's least squares solution for incomplete data. British Journal of Mathematical and Statistical Psychology, 1970, 23, 69-71.

Torgerson, W. S. Theory and methods of scaling. New York: Wiley, 1958.

APPENDIX B

LEARNING FROM TEXT: A SUMMARY OF NAVPERSRANDCEN RESEARCH

LEARNING FROM TEXT: A SUMMARY OF NAVPERSRANDCEN RESEARCH

John A. Ellis
Paula J. Konoske

Introduction and Background

An important problem in education and training is how to improve learning from text. Many educators and researchers believe that asking questions during learning is one way to accomplish this end. Anderson and Biddle (1975) provide a good summary review of studies designed to investigate the effects of practice or adjunct questions on learning. Some of the factors that have been studied include (1) question position (i.e., before or after the to-be-learned material), (2) type of questions (e.g., factual or inferential), (3) similarity of practice questions to the final test questions, (4) number of practice questions, and (5) frequency with which practice questions are given.

In the typical study, one or more groups given adjunct questions are compared to a group receiving no questions. Subjects in all groups are given a reading passage. In the adjunct question conditions, questions can be given in any of the ways described above. However, subjects are usually not allowed to refer back to the reading material to answer the adjunct questions. Rather, they must rely solely on memory. After completing the passage, all subjects are given a test to assess the effects of the adjunct questions.

Some of the more interesting findings occur in studies of adjunct questions that follow text passages (adjunct postquestions) as opposed to studies of questions that precede text passages (adjunct prequestions). Adjunct postquestions can have both direct and indirect effects. The direct effect is that postquestion groups perform better than a read-only control group on final test questions that are informationally similar or identical to the adjunct postquestions. The indirect effect is that subjects receiving postquestions perform better than control subjects on final test questions that are unrelated or incidental to the adjunct questions (Anderson & Biddle, 1975; Rothkopf & Bisbicos, 1967). The indirect effect is important because it shows that adjunct questions that follow sections of instructional materials can help the student learn information other than that covered in the questions.

Many investigators believe that the indirect effect occurs because the practice questions focus the student's attention on the type of question (e.g., factual, application) and/or type of information (main ideas or details such as dates, places, names, etc.) that will be included on the final test (Anderson & Biddle, 1975; Rothkopf, 1966). Focusing is thought to occur in two ways. First, the adjunct postquestions can alert the student to the type of information to study. This is called the "forward effect," and results in increased attention to the text following the questions (McGaw & Grotelueschen, 1972). Second, if it is assumed that students mentally review what they have just read in order to answer an adjunct postquestion, then they might also review, and perhaps learn, material similar in topic or near to the directly questioned material. This is called the "backward review" effect (McGaw & Grotelueschen, 1972; Rickards & DiVesta, 1974; Rothkopf & Billington, 1974).

The notion that questions focus attention follows from the assumption that students in a learning situation form hypotheses about what is important to study, and that adjunct postquestions facilitate this process by focusing students' study behaviors and information processing (c.f. Andre, 1979). Thus, adjunct postquestions have a positive effect because students use them to "figure out" what the final test will be like (Rothkopf, 1966; Rothkopf & Bisbicos, 1967).

The ideas that students actively attempt to discover what is important while reading text and that providing information about what to study enhances performance are supported by Brown, Campione, and Day (1981). They studied the effects of instructions and training designed to improve students' self-control and self-awareness of their own learning processes. They found that students learned more effectively when they were provided with instructions and training on how to study certain types of material and on how to monitor and control their study activities than when they did not receive instructions. Brown et al. (1981) concluded that, if learners can be made aware of basic strategies for reading and remembering, simple rules of text construction, differing demands of a variety of tests to which their information may be put, and the importance of activating any background knowledge which they may have, they cannot help but become more effective learners.

Further support for the hypothesis comes from Mayer and Bromage (1980) in their studies of "advance organizers." They found that a group that received a conceptual advance organizer before reading the text scored higher in recall of conceptual ideas, recalled more incidental material, and made more novel inferences on a recall test than did a group that received the advance organizer after reading the text. In terms of the present discussion, the advance organizer provided the students with "clues" about the nature of the final test and focused their study strategies and behaviors.

Navy Research

For several years, NAVPERSRANDCEN has been doing research on questioning and other related techniques for improving study effectiveness. This research has taken two, somewhat overlapping, directions. The first direction has involved working within the adjunct question paradigm described above to examine variables other than adjunct questions that control text processing. The second has examined the effects on learning and performance of manipulating practice variables in technical training programs. The remainder of this paper will briefly describe completed and ongoing research in each of these directions.

Adjunct Questions

The first study done in this area (Ellis, Main, & Ellis, 1976) was an attempt to extend Rothkopf and Bloom's (1970) work on oral adjunct questions. Four question conditions were run: a read-only control, a print-questions group, an audiotape questions group, and an in-person question group. The results of the posttest for adjunct questions were that all groups were superior to the control group and that the in-person group was superior to the other two experimental groups. However, because there were no differences among the groups for incidental test items, these results are inconclusive. Rothkopf and Bloom's (1970) finding that adjunct questions improve learning of incidental test items was not replicated.

Subsequent studies investigated another strategy for focusing student attention and study behavior. The strategy of providing adjunct postquestions was compared with the strategy of giving students instructions about the final test. Because it is assumed that the student is actively attempting to discover what is important to study, instructions about the test should facilitate learning. This facilitation should occur because instructions are "explicit," while adjunct postquestions are "implicit"; that is, the student must conclude what the test will be like from the type of questions asked during instruction. If the student forms a wrong hypothesis or concludes that only the information covered in the questions is important, performance on incidental questions may not improve. This is hypothesized to be the reason that incidental effects are not

found with adjunct prequestions (Anderson & Biddle, 1975). On the other hand, if the student can be given explicit information (or instructions) about what the test will be like before studying and about how best to "process" the information that is to be learned, test performance should improve. By "explicit," it is meant that the instructions are not vague or ambiguous and provide the student with essential information about the nature of the test and how best to process the information.

Duell (1974) provides some support for the notion that giving students information about what is important to study facilitates learning. She provided some students with detailed behavioral objectives which directed them to learn names and definitions in the text. Other students received a nonbehavioral objective that stated that they would be given a test after the reading. Following the reading of the text passage and the completion of the test, all subjects were asked to classify each test item as important or likely to be tested. Most of the items classified as unimportant tested recall of names and definitions. Duell found that students receiving detailed behavioral objectives performed better on these questions than did students given a nonbehavioral objective. She concluded that other strategies for focusing student study behavior may operate in a similar manner (i.e., by setting off information needed when the learner is tested and ensuring that the learner is aware of this relationship).

Further, Rothkopf (1966), in an investigation of adjunct questions, compared a group receiving specially prepared "care-inducing instructions" with adjunct question groups and a control group. The instructions included statements that the reading material contained much detailed factual information and that the text should be read carefully and slowly. The results indicated that instructions had general facilitative effects compared to effects on the adjunct questions and control groups. The results of this experiment support the notion that instructions can facilitate performance.

The first NAVPERSRANDCEN investigation in this area (Ellis, Montague, Wulfeck, Prince, Burnick, 1980) compared the effects of instructions and adjunct postquestions on learning verbatim factual information from text. The instructions were designed to be more explicit and detailed than Rothkopf's (1966) hortatory instructions. Students were given specific information regarding the nature of the test and the type of information to be tested, as well as general instructions about how to study the information to be learned. In this study, the questions were taken verbatim from the text.

The idea that giving students instruction about the final test can be as effective as giving them adjunct questions (and in some cases more so) in focusing attention and study behaviors was tested by comparing four groups: a read-only control group, an adjunct questions group, an instructions group, and a questions-plus-instructions group. The results showed that instructions can be as effective as adjunct questions in controlling text processing.

The next study (Van Beenen, Ellis, Wulfeck, Konoske, 1981) investigated the same question, but this time used paraphrased comprehension questions for both adjunct and test questions. The questions were constructed from a text on submarine warfare using guidelines given by Anderson (1972). The instruction group received information about how the questions were constructed and an example paragraph with comprehension questions. Results again indicated that instructions were as effective as adjunct questions in promoting general text comprehension, although directly questioned material was better learned in the adjunct postquestions conditions.

A third study (Ellis, Konoske, Wulfeck, Montague, 1982) also investigated the hypothesis that explicit instructions to students about how and what to learn can

facilitate learning of classification tasks as effectively as adjunct postquestions. Again, four groups of Navy enlisted personnel were tested: a read-only control, an adjunct questions group, an instructions group, and an adjunct-question-plus instructions group. The adjunct questions groups performed better than the control group on a test that required classification of both old and new instances of U.S. Navy call signs. Contrary to the previous finding, the results indicate that providing instructions about what a classification task is and how best to process the information was not as effective as providing adjunct questions and did not differ from the read-only treatment.

The instructional implications of these results are that students should be told explicitly what to expect on the test, as well as given adjunct questions. This strategy should allow students to concentrate more broadly on all information relevant to the final test rather than focusing on just the portion of the content covered by the adjunct questions. It should be noted that the strategy that would promote the most learning would be to include adjunct questions for every item to be tested; however, this approach is often impractical.

Practice Questions in Technical Training

A complementary line of research has been concerned with the effect of other types of practice on learning and performance in technical training and the relationships among practice and other variables that influence performance. While it has been well established that practice facilitates learning, the underlying mechanisms whereby this effect occurs are not well understood. We have hypothesized that practice affects learning in three general ways:

1. Practice focuses attention on specific tasks to be learned.

2. Practice informs the student about the nature and type of the criterion test for learning.

3. Practice (with feedback) allows the learner to monitor the progress of learning and perform self-diagnosis.

In addition, the inclusion of certain types of practice (e.g., recall) or instructions may have effects as in 1 through 3 above and therefore compensate for the lack of other types of practice (e.g., hands-on performance).

The first study (Merrill, Wood, Baker, Ellis, & Wulfeck, 1977) tested the notion that achievement on a final test will decrease if test items and presentation strategies are not consistent. That is, if the instruction contains all the information necessary to answer the test items but does not present this information in a manner that is consistent with answering the test items, test performance should suffer. Results indicated that, for both classification and recall tasks, a significant decrement in test performance and learning efficiency occurred when the lesson materials, including practice items, were not consistent. The results supported the hypothesis that instruction and practice inform the student about the nature and type of the criterion test.

A second study (Ellis, Wulfeck, & Montague, 1980) examined the effect of adjunct postquestions in an individualized training course. The training materials directed students to specific questions after they completed a portion of the instructional materials. Students could study them in any way they chose. To assess the effect on study behavior and learning, the percentage of adjunct questions identical to the lesson and final test questions was varied in four conditions: no questions, 0 percent identical,

50 percent identical, and 100 percent identical. The 100 percent group was superior on all measures. The remaining groups differed on several measures, including lesson and final test performance and performance on questions incidental to the adjunct questions. These results have direct implications for the design and development of individualized instructional materials. If the intent of the instruction is to have students remember important pieces of information (as in this study), then there should be a practice item and a test item for each piece of information. There should not be practice on information that will not be tested and there should not be testing on information that has not been practiced. If there are time and resource constraints so that not all the important information can be tested, the course material should be prioritized so that the most crucial information is practiced and tested. The same prescription would be made for classroom instruction where the teacher provides practice by oral quizzing or in workbooks. If the intent of the educational program is to have learning transfer to new situations, practice items would necessarily not be the same as test items, but both the instructional material and the test should provide an opportunity to practice and perform such transfer.

The next study involved two experiments and used material on U.S. Navy radio call signs to examine factors that affect learning of technical text. Ellis, Wulfeck, Montague, and Fredericks (1981), in the first experiment, examined the relationship between practice on to-be-learned material and instructions to the student about how and what to study. The results supported the hypothesis that practice focuses attention and informs the student about the nature and type of criterion test. Further, the results showed that instructions alone were as effective as practice alone; however, both together were the most effective.

The second experiment investigated the effects on learning two types of tasks of two types of practice: practice recalling conceptual characteristics or practice classifying instances of concepts. Findings indicated that practice in recalling facilitated both later recall and later classification, while practice in classifying only benefited later classification.

Additional experiments are planned to determine in more detail the effects of practice and instructions. The next studies will separate effects due to the act of practicing from those due to instructions or receiving feedback after practice.

The application of these findings to instructional design is apparent. The inclusion of practice recalling in classification tasks supports later recall and provides a sufficient basis for classification. Therefore, if time or cost constraints prohibit extensive practice classifying, practice recalling should be included, particularly in the absence of specific instructions to students about what and how to study.

Conclusions

Several general conclusions can be drawn from these two lines of research. Theoretically, the distinction has been made between implicit and explicit focusing of the learner's text processing. This had not been done before in the literature on questioning. The distinction is important for instructional designers because it provides a useful framework for thinking about the learner and adapting to his or her needs. We have demonstrated the powerful effects of instructions on processing relative to questioning. These results are especially interesting in view of early work by Anderson, Rothkopf, and others that failed to show an effect on learning from giving students objectives. The conclusion is that just giving students objectives is not enough. What is required is to give very explicit instructions about what and how to study. The results that have been

reported support this position. In addition, we have provided empirical support for the consistency hypothesis, and have shown that, when the standard adjunct question paradigm is adapted for use in a technical training course, the results are quite different from what happens in the laboratory.

****

Anderson, R. C. How to construct achievement tests to assess comprehension. Review of Educational Research, 1972, 42, 145-170.

Anderson, R. C., & Biddle, W. B. On asking people questions about what they are reading. In G. Bower, (Ed.). Psychology of Learning and Motivation (Vol. 9). New York: Academic Press, 1975.

Andre, T. Does answering higher-level questions while reading facilitate productive learning? Review of Educational Research, 1979, 49, 280-318.

Brown, A. L., Campione, J. C., & Day, D. Learning to learn: On training students to learn from texts. Educational Researcher, 1981, 10, 14-21.

Duell, O. K. Effect of type of objective, level of test questions, and the judged importance of tested materials upon posttest performance. Journal of Educational Psychology, 1974, 66, 225-232.

Ellis, J. A., Konoske, P. J., Wulfeck, W. H., & Montague, W. E. The effects of instructions and adjunct questions on classification task learning. Paper presented at American Educational Research Association, New York, 1982.

Ellis, J., Konoske, P. J., Wulfeck, W. H., & Montague, W. E. The comparative effects of adjunct questions and instructions on text processing. Journal of Educational Psychology, 1982 (in press).

Ellis, J. A., Main, R., & Ellis, K. The effect of adjunct question presentation mode on learning from prose (NPRDC Tech. Note 76-4). San Diego: Navy Personnel Research and Development Center, March 1976.

Ellis, J., Montague, W. E., Wulfeck, W., Prince, R., & Burnick, D. Controlling text processing: Adjunct questions vs. instructions. Paper presented at the American Educational Research Association Meeting, Boston, 1980.

Ellis, J., Wulfeck, W., & Montague, W. E. The effect of adjunct and test question similarity on study behavior and learning in a training course. American Educational Research Journal, 1980, 17, 449-457.

Ellis, J., Wulfeck, W. H., Montague, W. E., & Fredericks, P. The effects of instructions and practice on learning. Paper presented at the American Educational Research Association Meeting, Los Angeles, 1981.

Mayer, R. E., & Bromage, B. K. Different recall protocols for technical texts due to advance organizers. Journal of Educational Psychology, 1980, 72, 209-225.

McGaw, B., & Grotelueschen, A. Direction of the effect of questions in prose material. Journal of Educational Psychology, 1972, 63, 580-588.

Merrill, M. D., Wood, N., Baker, M., Ellis, J., & Wulfeck, W. Empirical validation of selected instructional strategy diagnostic profile prescriptions (NPRDC Tech. Rep. 77-43). San Diego: Navy Personnel Research and Development Center, September 1977. (AD-A045 309)

Rickards, J. P., & DiVesta, F. J. Type and frequency of questions in prose processing textual material. Journal of Educational Psychology, 1974, 66, 354-362.

Rothkopf, E. Z. Learning from written instructive materials: An exploration of the control of inspection behavior by test-like events. American Educational Research Journal, 1966, 3(4), 241-249.

Rothkopf, E. Z, & Billington, M. S. Indirect review and priming through questions. Journal of Educational Psychology, 1974, 66, 669-679.

Rothkopf, E. Z, & Bisbicos, E. Selective facilitative effects of interspersed questions in learning from written materials. Journal of Educational Psychology, 1967, 58, 56-61.

Rothkopf, E. Z, & Bloom, R. D. Effects of interpersonal interaction on the instructional value of adjunct questions in learning from written material. Journal of Educational Psychology, 1970, 61, 417-422.

Van Beenan, S. K., Ellis, J., Wulfeck, W. H., & Konoske, P. J. Controlling text comprehension: Text vs. instructions. Paper presented at the American Educational Research Association Meeting, Los Angeles, 1981.

APPENDIX C

INFORMATION PROCESSING CONSIDERATIONS FOR PERSONNEL SELECTION

INFORMATION PROCESSING CONSIDERATIONS FOR PERSONNEL SELECTION

Robert A. Wisher

The purpose of an aptitude test is to predict the probability of success in a learning or performance task. In the Navy, aptitude is measured by the Armed Services Vocational Aptitude Battery (ASVAB), whose scores are used to select personnel for school and job assignments. This traditional approach to the selection and classification of personnel, however, appears to have reached a plateau in its predictive power. These scores have failed to improve their predictive capacity; further, they have provided little data that can be used for tailoring the training for each student.

Recent work on human intelligence suggests an *entirely different approach* to the study of aptitude. This new approach emphasizes the dynamic, information-processing characteristics of an individual, rather than the static, item-response measure that has served as the standard among psychometricians. This Independent Research project attempted to develop alternatives to the traditional approach--alternatives that consider the dynamic nature of human thinking. A better understanding of the dynamics is expected to benefit significantly the Navy's efforts to select and classify its personnel as well as provide prescriptive information for individualizing training.

The ASVAB consists of 10 subtests that assess such aptitudes as word knowledge, arithmetic reasoning, and mechanical comprehension. The 3-hour, pencil-and-paper ASVAB examination is administered at the Military Enlistment Processing Stations (MEPS) (formerly called the Armed Forces Examining and Entrance Stations (AFEES)) and is scored on the basis of a simple item-response analysis of the multiple-choice items. The scores from the ASVAB subtests are combined in various ways in order to make decisions regarding the selection of personnel for various Navy ratings, or occupational specialties. Navy rating qualifications require, for example, a disbursing clerk to have a certain word knowledge and arithmetic reasoning total, whereas an aviation structural mechanic requires a certain word knowledge and mechanical comprehension composite. These cutoff scores are based on a statistical analysis between ASVAB scores and probability of academic success in the various Navy schools. However, schools and jobs require more than just knowing words and reasoning with numbers. Formulating hypotheses, keeping track of numerous alternatives, and reasoning in abstract terms, for instance, are important activities that an individual must perform in school and on the job. Such information-processing capabilities are not directly assessed by the ASVAB.

Information Processing

The information-processing approach to the study of intelligent behavior stems from the renewed interest in cognitive psychology and the increasing adoption by psychologists of models from computer science. Not satisfied with the long-standing stimulus-response approach to analyzing behavior, cognitive psychology attempts to unwrap the black box that accepts a stimulus as an input and emits a response as an output. Information-processing psychology is concerned with such constructs as short-term memory, processing speed and efficiency, and attentional limitations. Performance, then, is described as the operation of integral "programs" for dealing with information available from sensory channels and memory stores. The more efficient these operations, the better the performance. In the human context, operations that store, manipulate, and retrieve data have been called cognitive processes. Certainly, there are close analogies between information-processing models of a human and that of a computer; the psychologist, however, deals with a system that forgets, associates, and learns.

information-processing models are being applied to understanding the nature of intelligence in terms of dynamic real-time mental operations and are challenging the traditional approaches that represented human abilities, such as verbal comprehension, as uniform, statistical factors. Critics of the factor-analytic approach have expressed a need to understand the cognitive processes that are represented only by a summative score on intelligence tests. Since factor analysis is performed over a set of items, it is not possible to determine the processes employed in the solution of individual items, nor is it possible to determine in what ways individuals of varying abilities differ in their solutions of a single item, other than being less likely, for example, to answer correctly a multiple-choice item. The traditional approach, then, provides no means to identify directly the underlying processes that constitute intelligent behavior. The information-processing approach does not rely on the product measure alone but, instead, attempts to assess performance in terms of internal cognitive processes. By delineating how these processes function, one gains a more detailed understanding of the mechanics of intelligent behavior.

Two general approaches to investigating the cognitive processes involved in intellectual behavior have been referred to as the "cognitive correlates approach" and the "cognitive components approach." In each approach, behaviors are correlated with scores from ability tests such as the verbal skills score on the Scholastic Aptitude Test (SAT). The cognitive-correlates approach employs simple information-processing tasks. For example, the number of milliseconds it takes an individual to judge whether pairs of letters, such as AA or Aa, are physically-identical or name-identical can distinguish individuals with high verbal abilities from those with lower verbal abilities. This letter-matching task measures access time to highly overlearned verbal information, reflecting an individual's ability to recognize a visual pattern as part of a letter or word. Although this elementary information process takes only one-tenth of a second, its iterative use in verbal tasks makes it a significant factor in language-related skills.

The cognitive components approach, on the other hand, studies more complex information-processing tasks. A cognitive component is viewed as a collection of more basic cognitive processes that, when properly coordinated, constitute a discernable unit of thinking. Examples of experimental tasks considered more complex are the verbal analogy tasks (e.g., hand is to glove as foot is to ?) or linear syllogism task (Bill is taller than Paul. Paul is taller than Ed. Who is tallest?). Such tasks appear in reasoning ability tests; they encompass a more complicated set of cognitive processes that can be sequenced in various ways, depending on an individual's strategy or preferred "style." Clever experimental manipulations in displaying an item, such as a brief delay in presenting all the information, can affect performance in ways that are informative to researchers. In either approach, the rationale is that the individual differences observed in aptitude or abilities tests can be traced to individual differences in the speed and efficiency at executing the proper cognitive processes for task solution.

The traditional approach cannot deal adequately with the dynamic, cognitive processes engaged during task solution. Examining the step-by-step execution provides the type of process information that captures these dynamic activities. The emphasis of the present research, then, is not only on whether an individual solved a problem correctly but, rather, <u>how</u> the individual solved the problem.

Cognitive Components and Processing Efficiency

This study uses the cognitive components approach and attempts to assess performance on a reasoning task in terms of the efficiency with which the component processes

are executed. One problem inherent in this approach is in describing detailed information-processing activity in terms of rather loosely defined cognitive components. Our initial interest is not to provide hard and fast definitions of the component processes but, rather, to characterize individuals of varying abilities as being qualitatively different in the degree to which they employ the components in a systematic manner.

The programmatic, goal-oriented application of the component processes is central to our notion of processing efficiency. We use the term component in the manner described by Professor Robert Sternberg of Yale Univeristy, who proposed a new approach to research in problem solving. The basis of the framework, the component, was derived from a theory of human problem solving that depicts a set of elementary information processes operating upon internal representations of objects or symbols. According to Sternberg's proposal, three important properties associated with a component are (1) duration, or the real time expiring during the execution of a component, (2) difficulty, or the probability of a component being executed correctly, and (3) probabilty of execution. Sternberg further specifies functions that components perform, including (1) a meta-component that plans a course of action or algorithm, monitors the success of that algorithm, and decides alternative courses to take, (2) performance components that are responsible for direct execution of the algorithm, and (3) retention components that are used in retrieving knowledge regarding the problem-solving algorithm and intermediate products from this algorithm.

Eye Movements and Cognitive Processes

A major aspect of intelligence is the ability to solve problems. We have chosen eye-movement behavior to trace information-processing activities and thus trace the dynamics of problem solving. Eye movements are simply a visual record, called a scan pattern, of what an individual viewed in a display. A widely accepted view among researchers examining cognitive processes is that eye movements reflect the sequence, or stages, of problem-solving behaviors. This parallel was demonstrated in simple cognitive tasks involving mental rotation, sentence verification, and quantitative comparison. In a study that examined eye fixations during consumer decision making, researchers found that subjects used a pairwise comparison strategy in selecting their most preferred brand. This strategy was validated by retrospective explanations that were prompted by a slow-motion replay of the eye fixation sequence. Furthermore, the researchers concluded that the pairwise-comparison strategy was invoked primarily out of information-processing conveniences, and that subjects changed their strategies to adapt to different task demands. Importantly, eye fixation sequences were found to reflect the changes in strategy.

The present study addresses limitations of previous work and extends the eye scan approach to problems in personnel selection. Specifically, the study was undertaken to explicate eye scan components during solution of analogical reasoning problems. The frequency of occurrence, sequencing, and programmatic application of these scan components are of interest. Furthermore, we are interested in the extent to which individuals of varying ability, as measured by standard ability tests, exhibit different scan patterns. Insofar as eye fixation sequences reflect cognitive processes and scan components index their efficient application, individuals differing in aptitude measures should exhibit markedly different scan patterns.

Analogical Reasoning

Analogical reasoning refers to the process of drawing parallels from known information to obtain new information. Psychologists believe that analogical reasoning plays an important role in intellectual behavior. Examining the manner in which one plans and

executes an analogical reasoning task provides the type of process information that is necessary to step beyond the inherent limitations of the item-response approaches. Much of the figural analogy research, however, has been limited to the simple analysis of responses to items or subsets of items. These simple analyses cannot delineate the cognitive components involved in problem solving.

Figure C-1 provides an example of a figural analogy problem used in the present study. The upper portion, or the stimulus array, represents a 3x3 matrix of information, with the cellular entries varying progressively across a row or down a column. The task is to determine which of the eight distractor items, represented in the lower portion, or distractor field, fits the lower right cell of the matrix according to the progressive patterns. In the example provided, distractor #8 would be the correct answer.

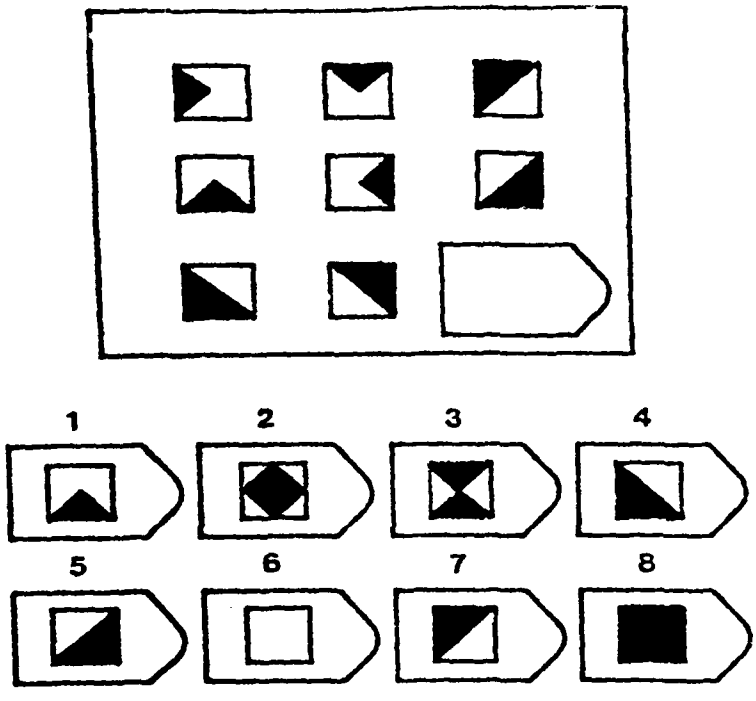


Figure C-1. Format of a typical analogy item.

The correct answer can be determined in a number of ways. For example, one might scan across the top row of the field, infer the transformation pattern, apply the rule across the bottom row of the field, induce the answer, and search the distractor array for the answer's corresponding number. This would be the ideal sequence of events. The multitude of possible transformations that must be considered within a row and cross-checked against the constraints imposed by column transformations, however, make the task nontrivial. Obviously, there are many cognitive processes engaged during such a complex reasoning task. The resourceful sequencing of the processes, however, can

reduce the amount of information that must be considered to arrive at the correct answer.

But what can eye scan patterns tell us about personnel selection? As will be seen from the description of the study, scan patterns can distinguish between average and low ability personnel as measured by the ASVAB, index the degree to which individuals prefer inductive versus deductive reasoning, determine the amount of information an individual needs before hypotheses can be formulated, measure how much information can be maintained in memory while performing associated tasks, and, in general, reveal a great deal about the mechanics of an individual's thinking. Importantly, this information can be obtained regardless of whether or not an individual "got the right answer" on an item. The extent to which these behaviors correspond to the intellectual demands of a school or job task will determine the extent to which eye scans can be valid predictors of school and job performance. The first step is to determine whether eye scan components can distinguish between ability levels and then test their predictive validity for school performance.

The Study

The purposes of the study were to develop a technique that captured the dynamics of problem solving, quantify the dynamics by means of "processing efficiency" indices, and test the power of the indices for predicting performance on an ability test.

Eleven Navy recruits were used in the study, six with above average ASVAB scores and five with scores at or below average. All subjects had normal vision without corrective lenses. Twelve analogy items were taken from the Ravens Advanced Progressive Matrices. The items were 3x3 figural analogies, as exemplified in Figure C-1.

An oculometer was used to track eye fixations. The oculometer is a device that computes the position of an eye fixation by monitoring corneal and pupil reflection of an infrared light. The reflection of light affects a voltage measure that, when properly calibrated, defines a horizontal and vertical line; the intersection of these lines defines the fixation point. Figure C-2 presents the system configuration. This technique provides reliable and detailed information on the sequence of eye fixations. By converting the analog signal to digital format, a permanent computer record of these fixation sequences is available.

based on pilot sessions conducted with a separate group of subjects, five scanning indices were devised. Each pattern represents one or more cognitive components, primarily combinations of the performance and retention components identified by Sternberg. To quantify each scan pattern, a scanning index was computed on the basis of eye fixation frequency. The five scanning indices were:

1. Percentage of the total number of fixations occurring in the stimulus array prior to the first excursion to the distractor field.

2. Number of excursions from a point in the stimulus array to a point in the distractor field.

3. Total number of fixations.

4. Ratio of total row scans (i.e., < 3 successive fixations between 3 consecutive points in a row) to binary row scans (i.e., > 3 successive fixations between 2 points in a row).

5. Ratio of total column scans to binary column scans.

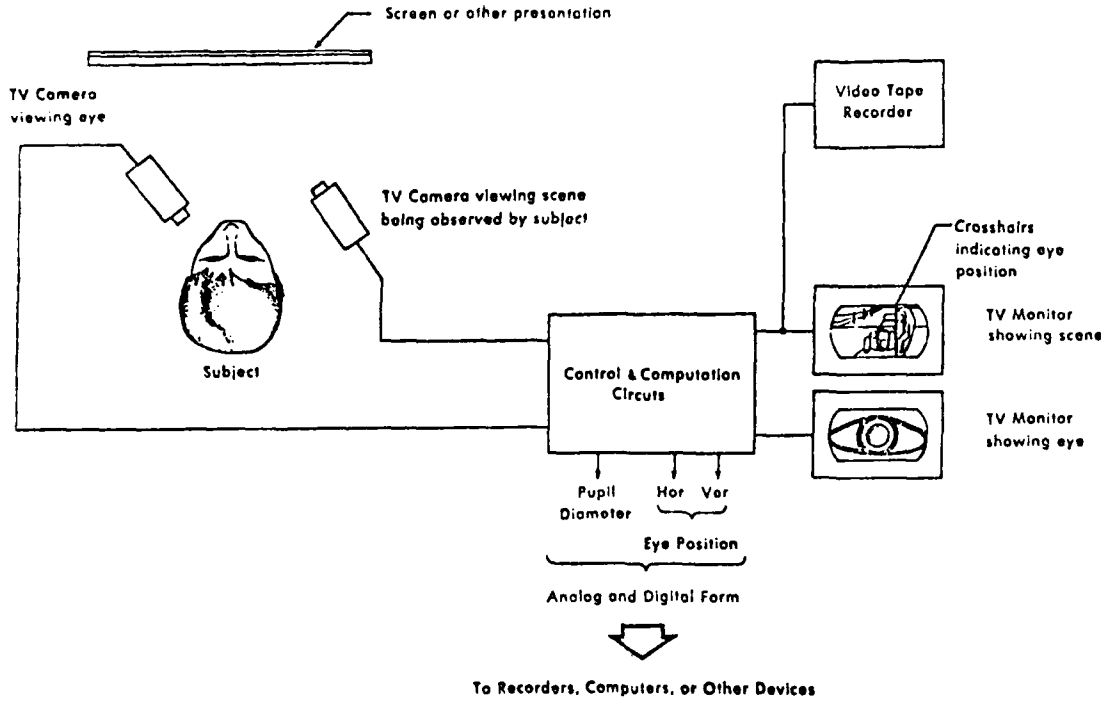


Figure C-2. System configuration.

These scanning indices were computed from the scan patterns of the experimental subjects. Correlational analyses were used to ascertain the relationship of the scanning indices to ASVAB performance. The correlations between the indices listed above and ASVAB total scores were -.70, -.73, -.59, .66, and .53 respectively. The five scanning indices thus constitute powerful predictors of ASVAB performance.

Examples of scanning patterns of high and low ability subjects are depicted in Figures C-3 and C-4 respectively. In these figures, each successive fixation over the 3x3 stimulus array is labeled and connected; excursions to the distractor field are truncated and labeled with a "d." In Figure C-3, the high ability subject exhibited tendencies for binary row scans, as in fixations 6-9 and 11-13. The low ability subject, Figure C-4, was less likely to exhibit this tendency; his scan patterns were more haphazard in contrast to the more designed, purposeful pattern apparent in Figure C-3.

Index 1, percentage of fixations occurring prior to the first excursion to the distractor field, reflects the extent to which examinees form complete hypotheses regarding the rules governing transformations between elements in the stimulus array. Individuals scoring high on the ASVAB were more likely to form a template of rules in the first and second rows of the stimulus array and to apply this template to the remaining portions of the array--to solve the item and then search the distractor field for the correct choice. Lower ability subjects engaged in a greater percentage of early excursions, indicating a preference for inductive reasoning. Higher ability subjects were more inclined to test hypotheses about single attributes or single transformations between stimulus elements.

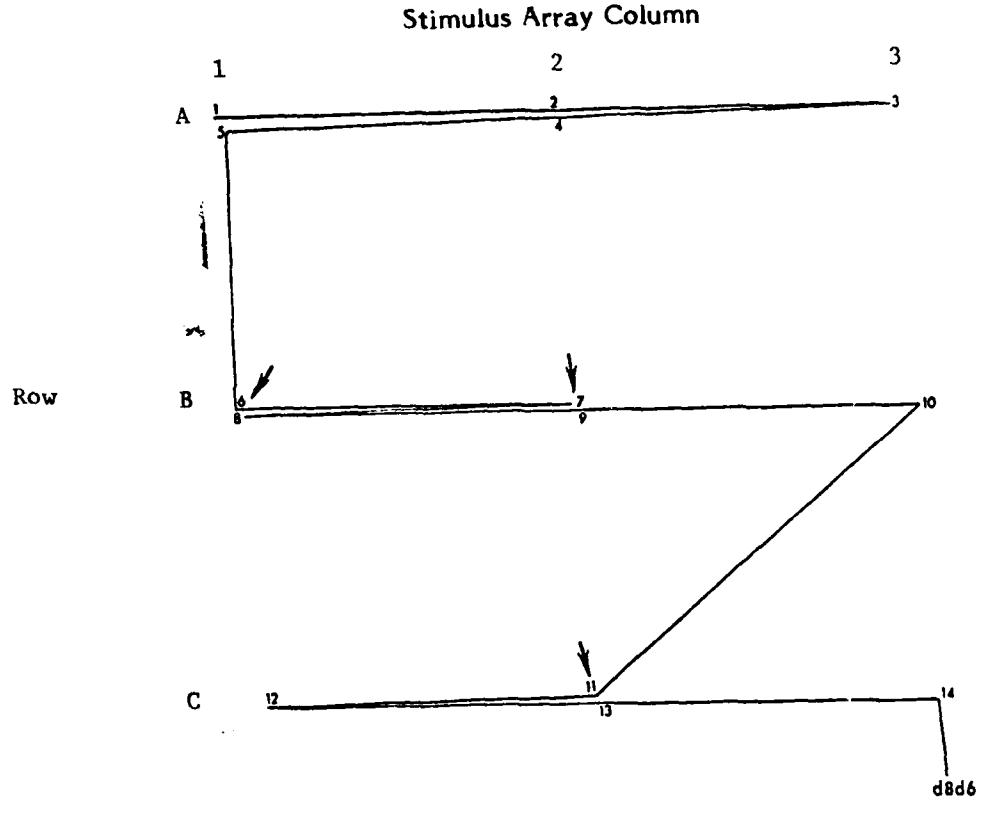


Figure C-3. Scanning sample of high ability subject.

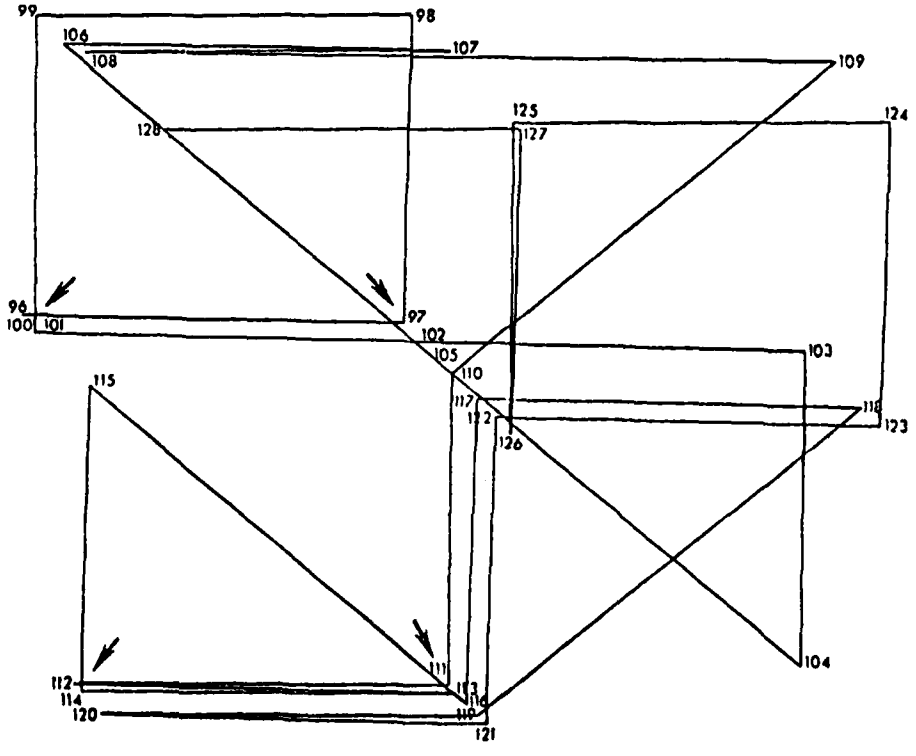


Figure C-4. Scanning sample of low ability subject.

Index 2, number of distractor excursions, reflects the frequency of moving from the stimulus array to the distractor field. Although information relevant to encoding of attributes, rule formation, and rule application may be derived by processing portions of the stimulus array, the distractor field provides only confirmatory information. The relationship of ASVAB performance to the point in intra-item solution at which the initial distractor excursion occurs indicates either confirmatory information or rule-relevant information is being sought from the distractors. Similarly, the number of distractor excursions reflects the frequency with which the individual abandons ongoing processing either to conduct a test or in an attempt to derive rule-relevant information outside the stimulus array. This scan pattern seems to indicate that frequent distractor excursions reflect the testing of single attributes or the trial-and-error search for something that "looks right."

A given number of fixations is necessary to derive the requisite information for item solution. Fixations in excess of this number provide no new information and thus are redundant. The finding that high ASVAB ability was negatively correlated with the number of fixations, Index 3, on the figural analogy reflects the tendency on the part of the lower ability subject to refixate a portion of the stimulus array, due either to inefficient processing or an inability to hold much information in memory while performing associated tasks.

The ratio of total row and total column scans to binary row and binary column scans, Indices 4 and 5 respectively, were both significantly related to ASVAB ability. The large versus small scan indices reflect "chunk" size and corroborate existing evidence that efficient processing of information is marked by an ability to deal with stimulus elements not as isolated bits of information but, rather, as meaningfully interrelated units. Individuals executing a relatively greater ratio of large scans perform better on the ASVAB due, perhaps, to an ability to integrate material into larger functional units and form hypotheses with less information.

Discussion

The concern of this Independent Research project is to consider the information-processing characteristics of individuals for purposes of personnel selection. Developing alternatives to the traditional approach to selection, which is limited to responses to test items, is necessary if the military is to improve its ability to select and classify its personnel.

This initial study forms the foundation for an approach that considers the information-processing characteristics of individuals engaged in intellectually demanding, problem-solving tasks. Efficient problem solving is viewed as the result of the programmatic use of a series of cognitive processes. These processes can be further collapsed into a set of cognitive components. The scanning indices developed here quantify the efficiency with which the components are used. These indices demonstrated high correlations with the ASVAB. The next step is to assess their predictive capability for school performance.

In addition to improving selection, this approach offers information that may be used for individualizing training. For example, the method can identify tendencies of individuals to perform a reasoning task inductively versus deductively or pinpoint weaknesses an individual has when formulating a hypothesis. Knowing these predispositions, an instructor is better able to explain a concept in terms that might be more comprehensible for one student but not necessarily so for another.

This study was a pilot; further research is underway to establish the robustness of the techniques. The technique has since been validated on a larger sample of college students. Interestingly, the indices were found to be better predictors of grade point average than the venerable SAT measure.

In summary, the advent of information-processing models of intelligence has clear potential for personnel selection. Information-processing models provide a better index of the effectiveness of thinking, whether during a test, in school, or on the job. As these models become more sophisticated and accurate, their validity as personnel selectors will grow.

APPENDIX D

PERSONNEL APPLICATIONS OF EVENT-RELATED BRAIN POTENTIALS

PERSONNEL APPLICATIONS OF EVENT-RELATED BRAIN POTENTIALS

G. W. Lewis

Introduction

The purpose of this paper is to summarize briefly NAVPERSRANDCEN research in the area of event-related brain potentials (ERPs) and their possible applications toward improving personnel assessment. For the last several years, NAVPERSRANDCEN has developed sophisticated laboratory capability in this area under in-house Independent Research and Independent Exploratory Development funding. Populations studied have included recruits, sonar trainees, and aviators. Current ERP research is directed toward improving security guard personnel assessment.

The military services depend heavily on paper-and-pencil testing to evaluate personnel. Results of such testing can be used to predict school and training performance reasonably well but are less effective in predicting on-job performance. The difference in predictive ability may relate to the way the brain processes information. The brain appears to have at least two different modes of cognitive processing. A verbal or analytic mode of processing information has been associated with left-hemisphere (LH) activity in most right-handed individuals; and spatial or integrative processing, with right-hemisphere (RH) activity. These two modes of cognitive processing were initially discovered by anatomical studies using as subjects "split-brain" patients suffering from accidental or surgical separation of the hemispheres. More recently, these processes have been assessed by measures of brain electrical activity such as electroencephalographic (EEG) and event-related brain potential (ERP) records, which show brain activity as very small signals recorded from the scalp. The EEG shows on-going activity, while the ERP shows activity after the brain has been stimulated (e.g., light flashes or clicks to the ears). For people performing verbal tasks, there is often decreased EEG/ERP amplitude over the left hemisphere. Spatial tasks often produce a decrease in amplitude over the right hemisphere, which may be considered an index of increased information processing within that hemisphere. Some individuals may use predominantly a verbal-analytic information processing style for learning, problem solving, and decision making, whereas others may use predominantly a spatial-integrative cognitive style for such tasks.

The reason that traditional paper-and-pencil aptitude tests predict academic performance but not on-job performance may be that such tests primarily tap the verbal, analytic processing performed by the left hemisphere. On-job performance may require much of the spatial or integrative processing performed by the RH. There have been many attempts to assess RH functioning by traditional testing procedures, with little success. Procedures like the ERP may not only tap RH processing to a greater degree, but predict on-job performance more accurately than the traditional paper-and-pencil tests. Assessing individual differences with an emphasis on "process" rather than "content" measures, employing the ERP as an index of brain activity, may prove more successful in assessing personnel and predicting human performance than the traditional tests. NAVPERSRANDCEN initiated the Applied Psychobiology Project in FY75 as the first major effort to apply the methods of psychophysiological research to the Navy's problems in the areas of personnel and training.

In NAVPERSRANDCEN, ERP measures have been shown to be useful in:

1. Predicting the success of Navy remedial trainees (Lewis, Rimland, & Callaway, 1976).

2. Describing the relationships between visual ERP measures and paper-and-pencil test scores (e.g., AFQT) (Lewis, [illegible]mland, & Callaway, 1977).

3. Differentiating pilots from radar intercept officers and relating brain ERP asymmetry relationships to aviator performance (Lewis, 1979a, b; Lewis & Rimland, 1979).

4. Showing visual ERPs as potentially useful in predicting performance of sonar operators (Lewis & Rimland, 1980).

5. Finding that ERP measures may be useful in problems relating to physical security personnel (Lewis, 1980, 1981).

More recently, it was shown that ERP brain wave measures may be related to methods of processing cognitive information (Lewis, Federico, Froning, & Calder, 1981). Also, a study was conducted to determine relationships between sensory interaction and reading ability (Lewis & Froning, 1981).

All of our studies have employed similar data collection procedures. Each subject was briefed on the research procedures and purposes. After the subject signed voluntary consent forms, he was prepared for recording. The technician cleansed the subject's hair and scalp at the electrode sites with an alcohol-impregnated cotton swab. A Lycra (elastic cloth) helmet was then placed on the subject's head. Recording electrodes were placed in contact with the scalp over predetermined (Figure D-1) sites.

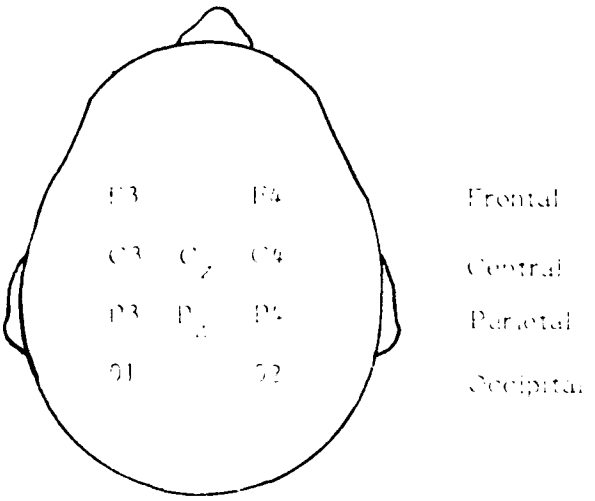


Figure D-1. Electrode site montage.

When the electrodes were in place and checked, the subject was instructed to observe his real-time EEG activity on the oscilloscope display. He was instructed to move his jaws and eyebrows so that he could observe how muscle artifact may contaminate the visual ERP (VERP) data. He was then seated in a darkened room in alignment with the visual stimulus and given a hand-held "time-out" switch that permitted him to suspend all stimulus presentation and analysis operations. He was instructed to press the switch to reject muscle artifact when he had to move.

The visual stimulus was a commercial fluorescent tube mounted in a 7" x 15" box with a power supply that was triggered by a computer-generated pulse. The stimulus duration was approximately 2 msec, flashed aperiodically, every 1 to 3 seconds. Stimulus intensity was moderate, at about 3 foot-Lamberts.

The EEG data were amplified, filtered, and then "averaged" in the laboratory computer to produce the VERP. The amplitude of each signal number was associated with each waveform. Figure D-2 shows typical amplitude data at sites corresponding to the locations in Figure D-1. The numbers displayed in Figure D-2 are μVrms times 100. The upper number at each site represents the μVrms value for the first 50 flashes, and the lower number, the second 50 flashes. The waveforms for each 50-flash series were superimposed on the same baseline. For each of the research projects reported in this paper, the first and second flash series were averaged.

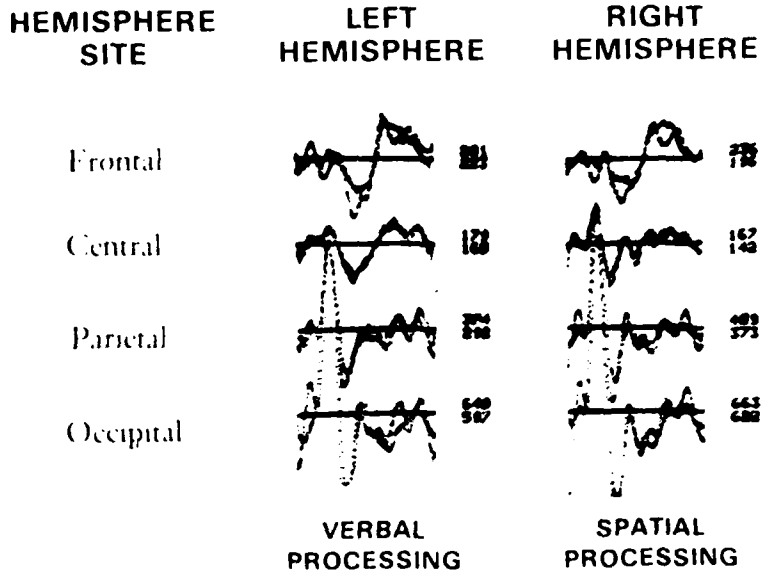


Figure D-2. Sample VERP data amplitudes (μVrms).

Data were obtained on a field-portable computer system (Figure D-3). The central processing unit was a Data General NOVA 2/10 equipped with a dual drive floppy disk unit, a small solid-state keyboard, an oscilloscope monitor, a fluorescent tube for visually stimulating the subject, and an integral eight-channel EEG unit.

Figure D-3. Subject wearing electrode helmet seated in front of computer system.

ERP Measures and Academic Performance

Three research projects dealt with ERPs and [illegible] related ERPs to predicting [illegible] second related ERP measure to [illegible] relationships between sensory [illegible]

A substantial percentage of the men recruited into the Navy are eventually found to be unsatisfactory. Many are discharged prematurely, and many of those who do complete their terms of enlistment fail to meet the standards for reenlistment. Our first research project (Lewis et al., 1976) was undertaken to evaluate newly developed computer-based methods of recording and analyzing VERP measures as a means of early identification of recruits with a high risk of premature discharge. The subjects (N = 73) were white males with an average age of about 19 years. Their scores on the Armed Forces Qualification Test (AFQT) fell between the 20th and 40th centiles, and they had scored between 3.0 and 5.5 grade levels on the Gates-MacGinitie Reading Test. As a result, the recruits were assigned to the Academic Remedial Training (ART) Unit, Naval Training Center, San Diego. Of the 73 men, 32 improved enough during ART to be continued on active duty (ACT group), while 41 failed ART and were discharged from the Navy (DIS group). The objective of this research was to find if the VERP measures could predict group membership.

A biserial correlation of .32 ($p < .05$) was obtained between frontal right hemisphere (RH) amplitude and the active duty-discharge criterion. By discriminant analysis, it was found that the same ERP amplitude measures contributed maximally to between-group variance: 62 percent of the subjects were correctly classified. Adding a second variate, parietal left hemisphere (LH) amplitude, improved prediction, increasing the number of correctly classified recruits to 68 percent (ACT = 56 percent, DIS = 76 percent). These results suggest that recruits prematurely discharged from recruit basic training tended to show VERP characteristics that distinguished them from men of approximately equal aptitude who successfully completed recruit training.

The next study (Lewis et al., 1977) related ERP measures to aptitude, as measured by the AFQT. The intent of this project was to determine if ERP data might augment information derived from the AFQT paper-and-pencil test for improved personnel classification and performance prediction. Subjects were 206 caucasian male Navy recruits (basic trainees), who had taken the AFQT prior to active duty. They were selected for inclusion in this study on the basis of their AFQT scores. The LOW group consisted of 103 recruits with AFQT scores ranging between the 20th and 40th centiles (an IQ range of about 87 to 96); and the HIGH group, of 103 recruits scoring between the 80th and 99th centiles (IQ range of about 113 to 133). The AFQT testing preceded the visual ERP testing by 7-15 weeks. Average age of the subjects was 19 years.

Statistical analysis (factor analysis and discriminant analysis) of the HIGH vs LOW aptitude group data showed significant relationships bertween the ERP measures and group membership. Sixty-four percent of the subjects in the verification sample were correctly classified by the ERP measures.

More recently, Lewis and Froning (1981) examined sensory interaction in a different sample of recruits. Visual (VERP), auditory (AERP), and bimodal (visual plus auditory, BERP) ERP measures were obtained from 41 subjects who were divided into two groups based on their reading ability (HIGH vs LOW). Sensory interaction was assessed by comparing responses to the bimodally presented stimuli with the responses to the visual and auditory stimuli presented individually. The greatest differences between the HIGH and LOW readers were obtained late in the ERP waveform, which suggested that sensory interaction greatly influenced high-order cognitive functioning. Reading ability is highly dependent on proper sensory interaction and higher cognitive functioning. It was suggested that distractability may partially account for the large ERP differences found for the two reading groups.

ERP Measures and Fleet Performance

Several research projects have addressed relationships between ERP measures and on-job, or fleet, performance. These projects have dealt with aviator and sonar operator performance, the prediction of enlisted promotions, and follow-up performance related to attrition.

Pilot and Radar Intercept Officer Performance

Naval aviators represent a highly selected and expensively trained group. Piloting a high performance aircraft would appear to require superior RH skills (spatial), while being a radar intercept officer (RIO) may place special demands on LH functioning (analytic). While each group must have above average ability in both the spatial and analytic areas, it nevertheless seems that the key elements of pilot and RIO performance might be categorized as primarily right- and left-hemispheric in nature, respectively. Could the pilots be discriminated from the RIOs based on RH and LH VERP amplitude measures?

To test this hypothesis, a field-portable computer system (Figure D-3) was mounted in the NAVPERSRANDCEN mobile laboratory facility to acquire VERP data from an F-4 fighter squadron of 58 aviators--28 pilots and 30 RIOs. There were three major objectives in this research: (1) determining the feasibility of recording data in an operational environment (these data are traditionally obtained in the laboratory), (2) seeing if we could determine individual and group differences in our subject sample based on the VERP data, and (3) relating performance of the aviators to our VERP amplitude and asymmetry measures (Lewis, 1979a, b; Lewis & Rimland, 1979).

Figure D-4 shows the laboratory van parked in the squadron hangar at the Miramar Naval Air Station, San Diego. Our first objective, recording the very minute VERP signals in this operational environment, proved feasible, even though there was a large amount of electrical and acoustical noise. The electrical noise was quieted by using special optical coupling in the VERP amplifier and filter system, together with a special electrical isolation transformer in the power line. Acoustical noise was reduced to an acceptable level by recording inside the van (designed with special noise-attenuating characteristics), and using white noise for masking.

Figure D-4. Mobile VERP laboratory parked on-site in squadron hangar.

VERP differences in the direction hypothesized were found between the pilot and RIO groups, accomplishing the second objective. These differences were greatest at the LH central and frontal sites. Because the selection criteria were very similar for both pilot and RIO groups, the differences between the groups could be attributable to training, to job requirements, or to experience. Further research would be required to clarify the finding.

This brings us to the third objective, that of seeing if the VERP measures from these aviators may be related to their individual performance levels. ERP hemisphere asymmetry relationships to performance have been of great interest to us. VERP asymmetry is an index of the differences between the voltages produced at analogous sites on the right and left sides of the scalp. The asymmetry value was defined as the RH amplitude minus that for the LH (RH - LH). Four asymmetry values were obtained simultaneously at the frontal, central, parietal, and occipital sites for both hemispheres. We have studied not only right versus left differences as they relate to performance, but also front-to-back relationships. The ERP asymmetry values for the frontal and central sites were averaged to provide the front measure, while the parietal and occipital asymmetry values were averaged to provide the back. Descriptive statistics (means and standard deviations (SDs)) were computed for each performance group. One of our most consistent findings shows that front and back VERP asymmetry SDs are related to aviator and sonar trainee performance levels, as well as to enlisted promotion rates. The SD is a measure of ERP dispersion and is one way of assessing individual differences in our personnel performance groups.

Individual and group asymmetry differences may be assessed at the same time by examining asymmetry SD values. Each of the pilots and RIOs were placed into high and low-performer groups based on flying proficiency, as judged by the squadron operations officer. A finding that has been consistent in the three projects described in this paper appears in Figure D-5. This figure shows the SDs plotted for groups (pilots and RIOs), performance ratings (high and low), and electrode sites (front and back). Left-handed and ambidextrous subjects were removed because hemisphericity often tends to be mixed in these subjects. The SDs for both the high-rated pilots and high-rated RIOs were about equal at the front and back sites. Also, for both high-rated groups, the SDs were greater for the back than for the front sites. The SDs obtained for the low-rated groups at the front and back sites were much greater than those for the corresponding high-rated groups. Further, the SDs obtained for low-rated pilots at the front and back sites were greater than those obtained for low-rated RIOs at these sites. As with the high-rated pilot and RIO groups, the SDs for the low-rated pilot and RIO groups were greater for the back than for the front sites.

The back electrode sites included a primary association area (parietal) and the primary visual reception area (occipital). The front site included both an association area (frontal) and a sensory-motor area (central). The task required only observing a blinking light; no muscle activity was needed. The greater heterogeneity of the low-rated pilot and RIO groups as compared to the high-rated groups may be a result of the fact that there may be many ways to perform poorly, but few ways to perform well.

A final observation in this project dealt with ERP habituation (Lewis, 1979b). Roughly half of the pilots and RIOs were instructors, while the other half were students. Visual ERP habituation was assessed by comparing the ERP records from the first 50 flashes with those from the second 50 flashes. The instructors showed visual ERP habituation while the students did not. This suggests that the instructors may have adapted more quickly to the experimental conditions and were less aroused than the students.

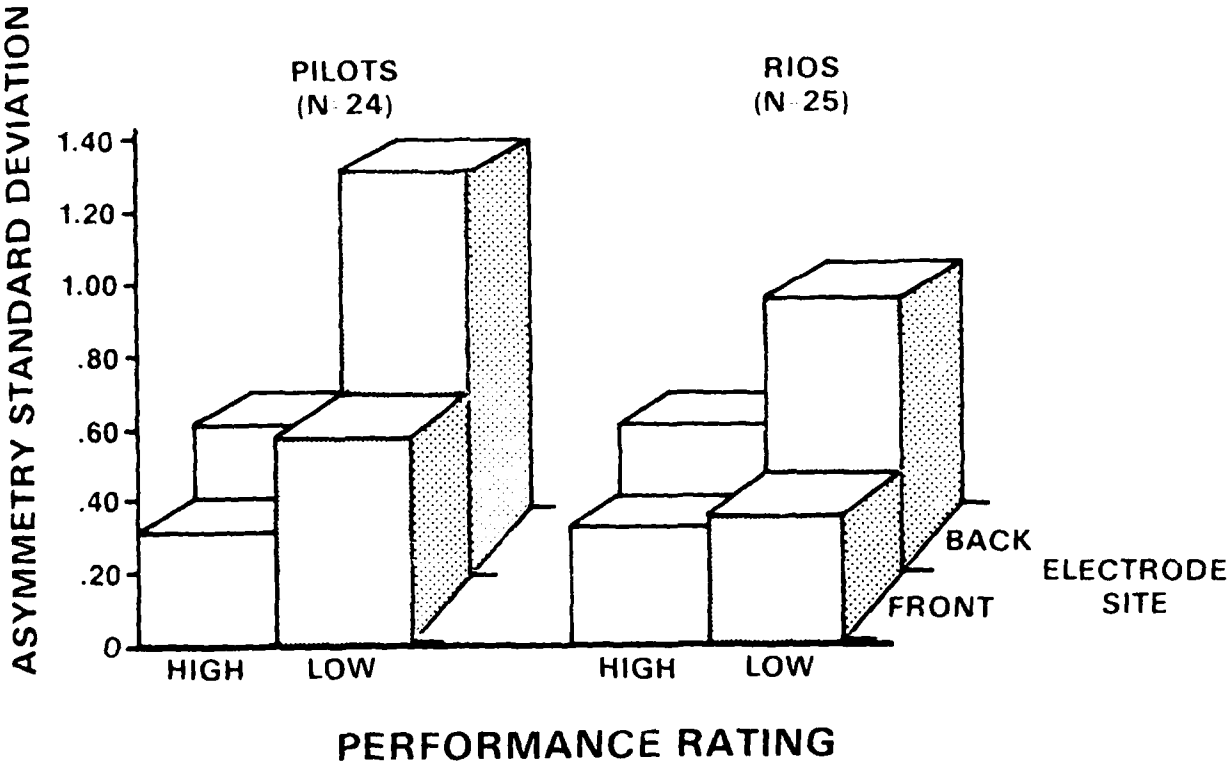


Figure D-5. Asymmetry standard deviations for the high- and low-rated pilots and RIOs, front and back electrode sites.

Simulator Performance of Sonar Operator Trainees

The operator of today's sophisticated sonar equipment must perform difficult and demanding mental operations requiring quick processing of visual and auditory information and the visualization of moving objects in three-dimensional space. Although conventional paper-and-pencil aptitude tests are reasonably effective in predicting academic performance in sonar school, they are not effective in identifying those who are most likely to perform successfully as sonar operators.

One objective in this research was to determine if VERP measures could be used to improve the prediction of performance of sonar operators (Lewis, Rimland, & Callaway, 1978; Lewis & Rimland, 1980). The sample of 26 ASW trainees was divided into two groups (HIGH and LOW) based on their performance on a sonar simulator. The HIGH (N = 14) and LOW (N = 12) groups showed no differences in their paper-and-pencil aptitude test scores (e.g., AFQT, arithmetic reasoning, mechanical ability) or classroom grade. However, substantial VERP amplitude (LH occipital site) differences were found.

Relationships between asymmetry and performance for the ASW trainees were similar to those for aviators, as may be seen in Figure D-6. Again we included only right-handed subjects (HIGH N = 10, LOW N = 10). The SDs, or asymmetry dispersion measures, were very similar from the front to the back of the head for the HIGH group. Greater

front-to-back differences were found for the LOW group than for the HIGH group. Perhaps the HIGH performers were able to integrate and use their entire brain more effectively than the LOW performers. Finally, there was less dispersion in both front and back regions for the HIGHs compared with the LOWs.

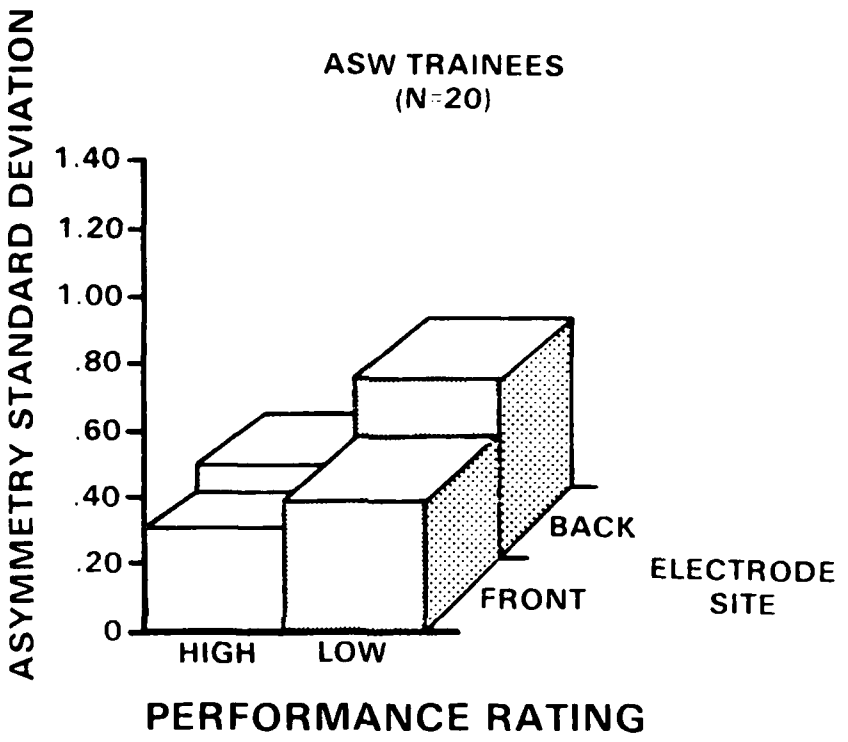


Figure D-6. Asymmetry standard deviations for high- and low-rated ASW trainees, front and back electrode sites.

Enlisted Promotion and Attrition

In a recently completed research project (Lewis, 1980), we obtained follow-up performance records for enlisted recruits 3 years after recording the initial VERP data. The primary objective was to compare the VERP amplitude and asymmetry predictors with the traditional paper-and-pencil aptitude and academic predictors used by the Navy. The subjects were the same as used by Lewis et al. (1976, 1977). Records from 37 recruits of the original sample of 279 could not be located after the 3-year period. Thus, the remaining sample (N = 252) was divided into two groups based on the number of promotions each enlistee achieved during the preceding 3 years. The HIGH group (N = 134) had two or more promotions, while the LOW group (N = 118) had less than two promotions. VERP amplitude measures were able to differentiate the two groups and classify (cross-validate) the subjects into the HIGH or LOW groups more effectively than did the traditional paper-and-pencil predictors, such as reading grade level or AFQT score.

Figure D-7 shows that the asymmetry dispersion relationships for the enlistees were similar to those for the aviator and sonar operator trainee groups. Unlike the aviator

and the sonar trainee groups, left-handed and ambidextrous subjects were included in Figure D-7. The front and back SD measures were lower for the HIGH than for the LOW performance group, as was found for the aviator and sonar groups. The HIGH group, however, showed a front-to-back difference that was not observed for the aviator or sonar trainee subjects. Back electrode site SDs were lower than the front SDs for the HIGH group enlistees, while the back SDs were slightly greater for the aviators and trainees.

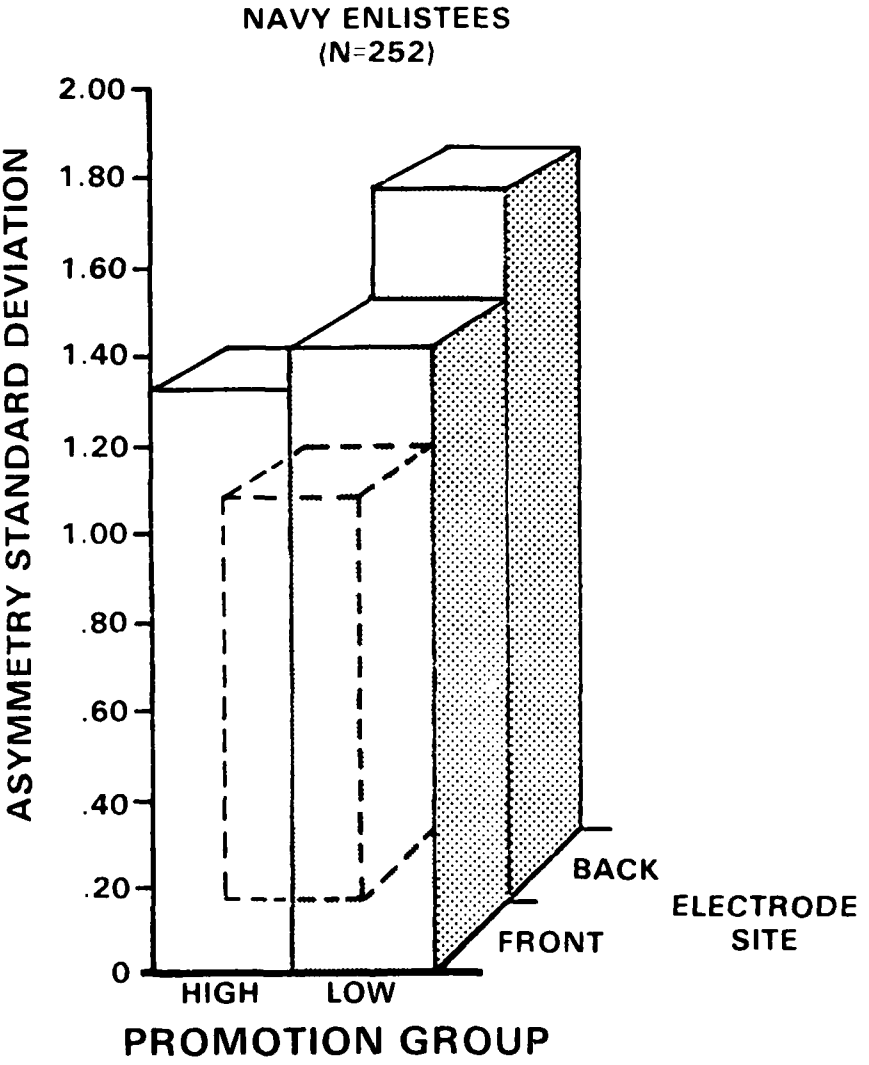


Figure D-7. Asymmetry standard deviations for high and low enlistee promotion groups, front and back electrode sites.

Attrition data were also obtained for the above sample of enlistees. Enlistees remaining on active duty (N = 143) were compared to those who were prematurely discharged from the Navy (N = 108). Of special interest were hemisphere information processing differences for these two groups. A large degree of brain asymmetry was found for the active duty group, but not for the discharge group. The ERP measures also

suggested that the active duty group was more homogeneous than the discharge group. Generally, the active duty group showed higher total promotions and lower desertion and absent-without-leave rates than did the discharge group.

Our most recent work has been directed toward determining the feasibility of applying ERP technology to training. This work involves assessing the unique capabilities of each individual in order to increase training efficiency. Initial results are promising. They show how critical the integration of the visual and auditory senses is to learning. Future research will deal primarily with assessing the feasibility of using psychophysiological measures to predict the tolerance of stress, evaluate performance reliability, and assess vigilance in security guard personnel.

NAVPERSRANDCEN has established an extensive library of ERP predictor and performance follow-up data from both the laboratory and operational environments. Research will continue to evaluate existing and new technologies for improved personnel assessment. One new technology under evaluation involves biomagnetic procedures (magnetic output of brain, muscle, and heart). Such procedures require superconducting/cryogenic equipment, utilizing liquid helium at -452°F. This new biomagnetic approach provides different information from the EEG/ERP procedures and has a decided advantage over EEG/ERP in that no direct contact with the body is required (Lewis, 1980, 1981).

Summary

Traditional paper-and-pencil aptitude tests may predict academic performance fairly well, but not on-job performance, because they tap the verbal, analytic processing primarily performed by the left hemisphere. On-job performance requires much of the spatial, simultaneous processing performed by the right hemisphere. There have been many attempts to assess RH functioning by traditional testing procedures, but with little success. Procedures like the VERP may, by tapping RH processing, permit us to predict on-job performance more accurately than can the traditional paper-and-pencil tests. Assessing individual differences with an emphasis on "process" rather than "content," as suggested by the concept of brain asymmetry, may prove more successful in predicting human performance than the traditional tests.

We have made a start in applying new advances in technology and new information on brain functions toward predicting and enhancing training and on-job performance. Several consistent findings have been observed in the various subject samples we have studied, especially the relationships between asymmetry SD measures and job performance. This suggests that ERP approaches may have widespread application to the general field of personnel assessment, training, and performance prediction. The new biomagnetic approach offers opportunities that may exceed those provided by ERP research.

****

Lewis, G. W. Field applications of evoked potentials. Presented to U.S. Air Force School of Aerospace Medicine (AFSC), Brooks Air Force Base, TX: 9 May 1979. (a)

Lewis, G. W. Visual event-related potentials of pilots and navigators. In Lehmann, D., & Callaway, E. Human evoked potentials: Applications and problems. New York: Plenum Press, 1979. (b)

Lewis, G. W. Job performance and brain asymmetry: Relevance for physical security personnel. Presented at the Fifth Annual Meeting on the Role of Behavioral Science in Physical Security, held 11-12 June at the National Bureau of Standards, Gaithersburg, Maryland: 1980.

Lewis, G. W. Biotechnology predictors of physical security personnel performance. Presented at the Sixth Annual Meeting on the Role of Behavior Science in Physical Security, held 3-4 June 1981, Washington, D.C.

Lewis, G. W., & Rimland, B. Hemispheric asymmetry as related to pilot and radar intercept officer performance (NPRDC Tech. Rep. 79-13). San Diego: Navy Personnel Research and Development Center, March 1979. (AD-A068 087)

Lewis, G. W., & Rimland, B. Psychobiological measures as predictors of sonar operator performance (NPRDC Tech. Rep. 80-26). San Diego: Navy Personnel Research and Development Center, May 1980. (AD-A085 030)

Lewis, G. W., & Froning, J. N. Sensory interaction, brain activity, and reading ability in young adults. International Journal of Neuroscience. 1981, 15, 129-140.

Lewis, G. W., Rimland, R., & Callaway, E. Psychobiological predictors of success in a Navy remedial reading program (NPRDC Tech. Rep. 77-13). San Diego: Navy Personnel Research and Development Center, December 1976. (AD-A037 339)

Lewis, G. W., Rimland, R., & Callaway, E. Psychobiological correlates of aptitude among Navy recruits (NPRDC Tech. Note 77-7). San Diego: Navy Personnel Research and Development Center, February 1977.

Lewis, G. W., Rimland, B., & Callaway, E. Visual event-related potentials: Toward predicting performance. In Callaway, E., Tueting, P., & Koslow, S. H. Event-related brain potentials in man. New York: Academic Press, 1978.

Lewis, G. W., Federico, P-A., Froning, J. N., & Calder, M. Event-related brain potentials and cognitive processing: Implications for Navy training (NPRDC Tech. Rep. 82-8). San Diego: Navy Personnel Research and Development Center, October 1981. (AD-A109 019)

APPENDIX E

SOME ISSUES RELATED TO THE IMPLEMENTATION AND EVALUATION OF OFFICE TECHNOLOGY

SOME ISSUES RELATED TO THE IMPLEMENTATION AND EVALUATION OF OFFICE TECHNOLOGY

John P. Sheposh
Vel N. Hulton

Background

It is becoming increasingly evident that we are in a transition from an "industrial" to an "information society" (Strassman, 1980). According to John Diebold (1981), Chairman of The Diebold Group, an international management consulting firm, we are on the threshold, by virtue of our increasing dependence on and use of computers, of becoming a "wired nation." The impact of these changes has been most pronounced in the office setting, where massive changes have already occurred due to computer technology. The period 1970 to 1977 saw as much change in the office as occurred during the previous century. It is estimated that the changes between 1977 and 1980 were as great as those in each of the two previous periods (The office-Vintage 197X, 1978). Correspondingly, it is predicted that the information-processing share of the U.S. economy will increase at a compound annual rate of 100 percent (Strassman, 1980). A case in point is the word processing industry, where the sales volume is expected to soar from 1.5 billion in 1978 to 6 billion in 1983 (Dunn, 1979). These trends, in addition to the spate of feature articles in popular magazines, the increasing coverage in trade journals, and the widespread offerings of symposia and workshops on white collar technology and its impact on organizational functioning, are indications of the significant changes that are taking place. The growth of office technology is not limited to the civilian setting; the same trends are in motion in the military.

The current level of activity and interest in office technology can be attributed to several factors. First, since the 1940s, the composition of the U.S. work force has shifted from being predominantly blue collar to white collar. By the year 2000, information workers will comprise over 60 percent of the working population, far outstripping the percentages for service occupations, agriculture, and manufacturing (Molitor, 1981). Moreover, over the last 10 years, white collar productivity has not kept pace with these other sectors. White collar productivity increased only 4 percent during this time period, whereas manufacturing and farming productivity increased 90 and 185 percent respectively. This relatively poor showing is due in part to the distribution of capital investment earmarked for technology for the various areas: The investment figures for the farm and manufacturing workers are $35,000-$50,000 and $25,000-$35,000 respectively, compared to only $2,000-$3,000 for the office employee (The office-Vintage 197X, 1978; Kurshan, 1981; Now the office of tomorrow, 1980; Schatz, 1981). Because of the tremendous drop in the cost of computer hardware components and steep rise in labor costs, it has been proposed that greater capital investment in office technology will provide a cost-effective response to the problem of diminishing productivity. It is estimated, for example, that, without office technology, white collar direct costs could rise to $1.5 trillion by 1990. With office automation, however, the costs would be reduced by more than $300 billion annually by the end of the next decade (Poppel, 1981).

These comments have pointed to signs of a growing consciousness of the potential of office technology as a way to enhance white collar productivity. The dramatic developments in office technology and its growing acceptance underline the importance of addressing the issues and problems associated with the implementation and evaluation of these systems. It is especially important to recognize the potential that office technology possesses for radically changing the shape of work in the office environment. The thrust

of this article is to call attention to some of the implementation and evaluation issues that are likely to arise as these new technological systems are integrated into the flow of work.

Implementation and Evaluation Issues

The importance of implementation and evaluation issues to management and evaluation specialists alike cannot be overemphasized. As Eveland (1981) observes in his discussion of the implementation of technology in the organization: "The manner in which a complex innovation is implemented may account for more of the variance in the net benefits realized than do its main technological features per se" (p. 2). In order to fully appreciate the issues that arise when these new systems are implemented, it is useful to consider the term "implementation" as it is used. Implementation refers to the transformation of a proposed change into practice. It is at this juncture of the change process that the proposed change confronts the reality of the organizational setting. Successful implementation implies an enhanced ability of an organization to accomplish its assigned tasks as facilitated by the innovation. In addition to understanding the concept of implementation, it is also important to recognize the relationship of evaluation to the implementation process. Evaluation is an integral part of the implementation process since it measures what is to be achieved, how it is achieved, and the extent to which it is achieved.

Prospective Users' Views of Office Technology

Successful implementation of change depends largely on the extent to which prospective users understand the scope, function, and impact of the change. The introduction of technology to the office setting is particularly susceptible to problems arising from lack of clarity concerning these change-related aspects. The term office technology or automation seems to evoke different images in prospective users. In a recent investigation (Sheposh & Hulton, 1981a), in which we interviewed a group of professionals and managers from a public sector organization as to their concerns about a new office technology system, three distinct perceptions of the technology emerged. Several viewed the technology as impacting solely on support personnel. Others saw it as a tool for managers and professionals as well as support staff that could be employed in specific areas of their work. Finally, there were those who viewed the technology in terms of an integrated system that would aid management to execute more effectively its mission and achieve its goals. These differing interpretations very probably have a direct influence on the prospective users' perceived role with respect to the technology and, in a less direct fashion, in their commitment to it.

These diverse reactions can be attributed, in part, to the differences in familiarity with automated office equipment and in the duties and responsibilities confronting each respondent. It is, therefore, not surprising that, when office technology is introduced, there is an absence of consensus regarding the purpose of the innovation, a lack of clarity of the users' exact roles, a lack of agreement as to the required skills, knowledge, and materials, and varying intensities of motivation and commitment. This makes the difficult task of implementation even more difficult.

The fundamental objective of office technology is an increase in productivity. Most of the productivity gains have been realized at the support levels of the organization, such as that obtained through the use of "word processing." This has fostered the common idea that office automation is applicable primarily to certain well defined functions at certain levels of the organization. The broader capabilities of word processing and the emergence of other technological possibilities in the office automation field today are such that they

are capable of providing professionals and managers, as well as the support staff, with the means to perform their job functions better.

Anderson (cited by Gottheimer, 1981) contends that the concept of office automation is restricted and outdated and feels that the reason that the term is still used is that the technology is being implemented from the bottom up. He notes that there are three general levels of management in every organization: (1) specialists and analysts who manage themselves, (2) managers and supervisors who manage people, and (3) executives who manage assets and risks. He states: "So far, the hardware and software have only barely gotten to the first layer. When it reaches the third level, bosses will start calling it management technology" (p. 15). To fully appreciate the capability of these systems, the perception must change from one of systems that simply "do things" to one of systems supporting all individuals in accomplishing their work.

Clearly, the scope of change, as defined by the users and those responsible for the implementation, is of central importance in any implementation effort. The technology is more likely to receive the necessary commitment, both attitudinally and in terms of the provision of time, personnel, and resources, if it is seen as serving a variety of functions with the potential of benefitting individuals at all levels of the organization. To summarize, the perceived scope of change determines the leverage for change, provides the structure for the kinds and ways in which information concerning the system is provided, and, finally, helps determine the nature of the criteria necessary in evaluating the impact of this new technology.

Selection of Outcome Measures

The different ways of conceptualizing technology in the early stages of implementation have a bearing on the choice of outcome measures or criteria selected, since they are linked to the perceived functions of the system. In general, outcome measures are seen in terms of the productivity and effectiveness of the organization. Probably the most prevalent view is the production orientation that focuses on savings at the support level. Since the variety of work required in the office often goes beyond standard production work and since the new systems can facilitate such work by providing automated assistance to managers and professionals, the prevailing focus on the support level is limited.

In some situations, the nature of the work suggests a specific outcome measure. However, in situations where the planned innovations do not dictate specific outcomes, it is more difficult to achieve exact specification of outcome variables. Outcomes may range from "hard" economic variables (e.g., units produced per work hour) to "soft" social psychological variables (e.g., organizational commitment, group cohesiveness). Some suggested economic indices measure labor efficiency. They are defined in terms of time (total staff hours to total volume) and costs (total salary costs). Equipment use is another index defined with respect to time (total time used to total quantity of equipment) and costs (total equipment costs) (Stout, 1981). The difficult task of measuring managerial productivity has been tackled by Williams (1980). Through several methods, the amount of time managers spend on inefficient practices such as queuing at the copying machine and delivering telephone messages is determined. The amount of misallocated work is then translated into cost figures. Thus, management inefficiency is quantified, at least to some degree.

Further, the various economic and social psychological variables may be combined in some fashion to effect overall organizational effectiveness (cf. Kilmann & Herden, 1976).

To assess the impact of technological systems in a meaningful way, an approach that incorporates a variety of outcome measures is recommended, particularly in those situations where the change is not completely predictable or the effects of the innovation have broad ramifications throughout the organizaion. Furthermore, assuming that multiple measures more closely approximate the construct of overall organizational effectiveness, emphasis on only one index may be misleading.

While it makes eminently good sense to select outcome measures first and then select the technologies that will help attain the desired ends, it is often not that clearcut in the day-to-day operations of a working environment. At times, because of cost concerns or deficiency in technological advancement, the necessary hardware is not available. Also, some organizations are either reticent or incapable of articulating their objectives (Sachs, 1981).

To ascertain the impact of newly introduced technological systems on organizational effectiveness, it is also necessary to obtain measures that tap the implementation process. The need for implementation information or short-term feedback cannot be overemphasized. Since, in the process of implementing an innovation, unanticipated problems will probably be encountered, feedback bearing on the progress of the implementation is essential in guiding and modifying, if necessary, the implementation plan. In particular, it is necessary to determine whether the technological system is actually in use and to what extent. While this may seem patently obvious, Hall and Loucks (1977) point out that "the assumption that the experimental group members do, in fact, use the 'innovation' and that the comparison group members do not, needs to be addressed systematically rather than left to chance" (p. 264). A case in point is Berman and McLaughlin's (1978) description of innovation "cooptation" as the tendency to attend selectively and exclusively to those features of the innovation consistent with existing practices. The importance of first-hand knowledge concerning the level of use of the equipment is indispensable in the determination of its impact, since this information helps to militate against false assumptions and misleading interpretations about the effects of the innovation.

In sum, the configuration of new office technology systems is increasingly aimed at significantly improving organizational effectiveness as a whole. As a result, measurement of a specific change in an isolated instance cannot adequately assess the overall level of effectiveness. This complicates the selection of outcome measures. The task demands thorough knowledge of the systems, the structure and function of the organization, its management, and the nature of the workforce. Because offices may vary along these dimensions, the selection of indicators of effectiveness and the assigned weighting of their relative importance must necessarily be tailored to the specific organization.

Prospective Users' Resistance

The third issue in the implementation of office technology is the resistence of prospective users to the new systems. These highly complex technological systems have the potential for producing stress and possibly resistance in two ways:

1. The technology has advanced to the stage where the way work is conducted is fundamentally changed. For example, it is possible for managers to accomplish their work in paperless, deskless offices.

2. Interacting with the technological hardware may also be psychologically stressful and inhibiting (Elam, 1980). The user at an on-line terminal may feel intimidated by the immediate response of the system and may feel pressured to respond hastily. In

addition, the user may view the computer as remote, self-governing, and unfriendly, and may react by feeling unsympathetic toward the system.

Thus, the potential for stress and accompanying resistance to the equipment must be examined (Klee, Sojka, & Crisp, 1981).

Certainly resistance, as expressed by users' attitudes, may have a bearing on their acceptance of the systems. For example, Abrams, Sheposh, Cohen, and Young (1977) assessed the attitudes of sonar operators towards a new sonar system and towards change in general. The results supported the hypothesis that a relationship exists between performance and system-specific attitudes; the higher the performance level on the system, the more positive the orientation towards the system. A partial explanation for this relationship is that the change blocks satisfaction of certain employee needs, thus creating negative attitudes that result in resistance (Mealiea, 1978). It is our view, however, that the widely held assumption that the fate of planned organizational change is largely or entirely a function of the initial resistance to the innovation by the member of the adopting unit is overextended. Elam (1980) reports that most new users have a provisional, noncommittal attitude toward the new system. In our previously mentioned study (Sheposh & Hulton, 1981a), attitudes indicative of resistance to the system were not evident at the time of implementation of the system. In general, office support personnel were in favor of the new system. Of the 20 people interviewed, 12 were positive or very positive, 6 were neutral, and 2 were negative. A similar pattern was obtained for managers and professionals. The majority were positively disposed to the new system. Five reported intense interest in using the equipment, six reported moderate interest, and three were either neutral or indifferent toward the system. Overall, the proposed office automation was welcomed as an opportunity to do something new and different, and its advantages outstripped any disadvantage. Similar results were obtained in other more recent investigations of implementation of office technology systems: (1) the productivity enhancement project (PEP) at the Naval Surface Weapons Center, Dahlgren, VA, and (2) the microfiche image transmission system (MITS), Navy Annex, Washington, D.C. (Sheposh & Hulton, 1981b).

Thus, the fact that attitudes have been found to influence the degree of acceptance of an innovation should not be automatically viewed as an indication of an individual's initial resistance to change. As Gross, Giacquinta, and Bernstein (1971), in their intensive study of an educational innovation, argue, resistance can develop after an innovation has been introduced; that is, during the period when implementation is attempted. Significant attitude change may occur as a function of the experiences one has with the equipment. Thus, the way change is implemented (e.g., the kind and amount of information about one's role vis-a-vis the innovation) may have a greater bearing than initial resistance on the development of resistence to the technology that is so often cited as a primary reason for the less than optimal use of the equipment.

Proposed Schematic Overview

Raymond P. Kurshan, President and Chairman of Management Assistance, Inc., a multinational company that markets and services word processing equipment, stated in an address to the Rotary Club of New York (1981):

> When properly implemented.... technology does provide greater control over time and activity. It provides needed information when it is needed and in a usable form ... gives managers greater

> flexibility and control in setting priorities and evaluating performance. I preceded the list of what office technology does and does not do with the caveat: "When properly implemented." (p. 21).

This statement emphasizes and resonates with the major theme of this article, which is that implementation efforts are central to the success of office technology in improving organizational effectiveness.

The present article has reviewed some of the major issues in the implementation and evaluation of organizational change in general and office technology in particular. The question then arises: What are the components or strategies that constitute an effective implementation? Some of the major elements that must be incorporated into an effective implementation program are presented in Figure E-1.

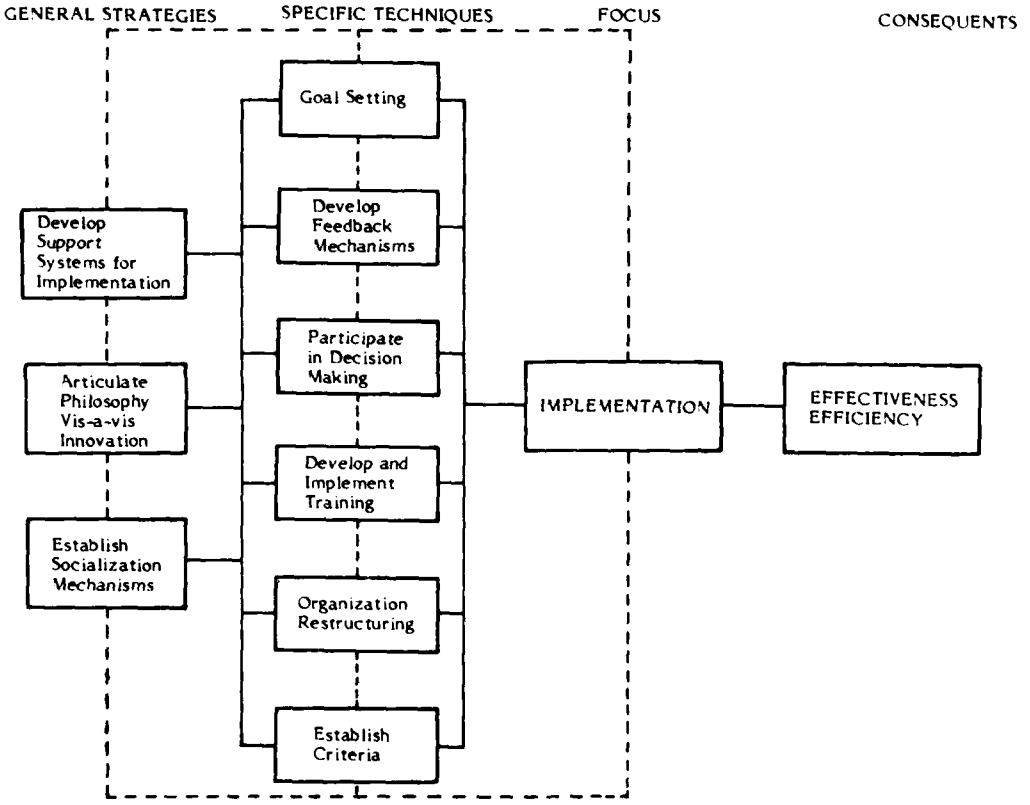


Figure E-1. Elements of an effective implementation program.

Inherent in this schematic overview is the assumption that the extent to which office technology increases organizational effectiveness depends on the kind and amount of attention focused on the implementation process and issues related to it. The focus of implementaion as shown in Figure E-1 is manifested in three interrelated strategies. First, there must be an effective support system for implementation of technology. This would include such specifics as provision of funds, training materials, rearrangement of prevailing organizational structure so that it is compatible with the function and goals of

ıvation, and availability of local resource personnel to provide practical advice. an organizational philosophy concerning the technology must be articulated that the scope of change, the function it serves, its impact on people at each level of ınization, and its impact on formal and informal organizational structures. Third, ıtion mechanisms must be established in which attention is given to reducing the ween the employees' perceptions of the innovation and the views articulated by the ıtion. This might invoke such specific techniques as emphasis on personal benefits nnovation and creation of pockets of commitment. As can be seen in Figure E-1, eneral strategies are translatable into specific managerial techniques, which are ry if the general strategies are to be effected. Thus, the implementation effort nes the extent to which the technology is integrated into the organization and ely determines organizational effectiveness.

Y

e white collar work force is growing rapidly. Office costs are doubling roughly 0 years and the need for increased productivity and efficiency is increasingly . Technological breakthroughs over the past few years appear to be part of the The benefits of these technological systems can be realized, however, only if e properly implemented and evaluated. The central issues bearing on these es have been addressed in order to provide a framework suggestive of some of the strategies and specific techniques that might be employed to effect successful entation.

****

, M. L., Sheposh, J. P., Cohen, P. A., & Young, L. E. Sonar operators' attitudes beliefs: Effects on introduction of new systems. (NPRDC Tech. Rep. 77-18). San o: Navy Personnel Research and Development Center, February 1977. (AD-A036

, P., & McLaughlin, M.W. Federal programs supporting educational change, Vol. Implementing and sustaining innovations. Santa Monica: Rand, May 1978.

, J. Increasing office productivity through information technology. In J. M. w (Ed.), Productivity: Prospects for growth. New York: Van Nostrand, 1981.

. The office of the future - Part I. Computer Decisions, July 1979, 16-19; 23; 26.

. G. Human considerations. In Depth, 1980, 2-10.

, J. D. Evaluating the implementation of organizational technology. Paper nted at the Annual Meeting of the Evaluation Research Society, Austin, Texas, ber 1981.

N., Giacquinta, J. B., & Bernstein, M. Implementing organizational innovations: A logical analysis of planned educational change. New York: Basic Books, 1971.

imer, D. Advanced office automation moving up to manager ranks. Duns Review, h 1981, 117(3), 14-16; 20.

Hall, G. E., & Loucks, S. F. A developmental model for determining whether the treatment is actually implemented. American Educational Research Journal, Summer 1977, 14(3), 263-276.

Kilmann, R. H., & Herden, R. P. Towards a systemic methodology for evaluating the impact of interventions on organizational effectiveness. Academy of Management Review, July 1976, 87-98.

Klee, K., Sojka, D., & Crisp, W. Changing the shape of work. Datamation, August 1981, 28-31;34.

Kurshan, R. P. White-collar productivity. The Rotarian, 1981.

Mealiea, L. W. Learned behavior: The key to understanding and preventing employee resistance to change. Group and Organizational Studies, 1978, 3(2), 211-223.

Molitor, G. T. T. The information society: The path to post-industrial growth. The Futurist, April 1981, 15-(2), 23-30.

Now the office of tomorrow, Time, November 17, 1980, 80-82.

Poppel, H. L. The automated office moves in. Datamation, November 1981, 25(13), 72-77.

Sachs, R. T. Matching the technology to the organization. Administrative Management, June 1981, 42(6), 36-39; 68.

Schatz, W. Advice for Uncle Sam. Datamation, January 1981, 67-68.

Sheposh, J. P., & Hulton, V. N. Summary of interview data on office technology project. Unpublished manuscript, Navy Personnel Research and Development Center, 1981. (a)

Sheposh, J. P., & Hulton, V. N. Studies of implementation strategies for managing change. (Work in progress) 1981. (b)

Stout, E. M. Notes on the current state-of-the-art for measuring and evaluating advanced office systems, and why an expanded approach is needed. (Working paper). Unpublished manuscript, U.S. Department of Housing and Urban Development, 18 September 1981.

Strassman, P. A. The office of the future: Information management for the new age. Technology Review, December/January 1980.

The office - Vintage 197X. The Current Office, June 1978, 101-109.

Williams, L. Office automation, organizational change and management productivity. Presented at Academy of Management Meetings, Detroit, August 1980.

DISTRIBUTION LIST

Chief of Naval Operations (OP-115) (2)
Chief of Naval Material (NMAT 00), (NMAT 04), (NMAT 05) (3), (NMAT 08)
Chief of Naval Research (Code 400) (15)
Chief of Naval Education and Training (02), (022), (N-5)
Chief of Naval Technical Training (Code 016)
Commander Naval Air Systems Command (NAIR-03) (4)
Commander Naval Electronic Systems Command (NELEX-06)
Commander Naval Facilities Engineering Command (NFAC-03) (2)
Commander Naval Sea Systems Command (NSEA-003) (5)
Commander Naval Surface Weapons Center
Commander Naval Air Development Center
Commander David W. Taylor Naval Ship Research and Development Center
Commander Naval Weapons Center
Commanding Officer, Naval Coastal Systems Center
Commanding Officer, Naval Underwater Systems Center
Defense Technical Information Center (DDA) (12)


FILME
8-8